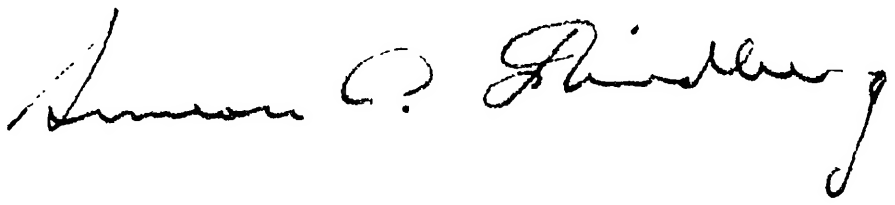

THEORETICAL PHYSICS

THEORETICAL PHYSICS

BY

W. WILSON, F.R.S.

VOL. I. MECHANICS AND HEAT
NEWTON—CARNOT

VOL. II. ELECTROMAGNETISM AND OPTICS
MAXWELL—LORENTZ

VOL. III. RELATIVITY AND QUANTUM DYNAMICS
EINSTEIN—PLANCK

THEORETICAL PHYSICS

by

W. WILSON, F.R.S.

HILDRED CARLILE PROFESSOR OF PHYSICS
IN THE UNIVERSITY OF LONDON,
BEDFORD COLLEGE

VOL. III
RELATIVITY AND QUANTUM DYNAMICS

EINSTEIN—PLANCK

WITH 23 DIAGRAMS

METHUEN & CO. LTD., LONDON

36 Essex Street, Strand, W.C.2

First published in 1940

PRINTED IN GREAT BRITAIN

PREFACE

THE same guiding principles have been followed in this concluding volume as in its predecessors, the chief one being that of presenting physical theory as a coherent logical unity. It has been necessary of course to omit many things, even important things, from the work ; but I feel that little has been omitted which is important from the point of view of the principle mentioned.

Relativity, which occupies the earlier part of the volume, has been introduced in a way which I believe to be new. Einstein's special theory of relativity is developed in what seems to be the most natural way, i.e. by following the suggestions contained in the 4-dimensional character of many sets of equations contained in both of the Volumes I and II, more especially those contained in the Maxwellian field equations as they appear on page 114 of Volume II. Moreover, during the development of the special theory the general theory has been kept in mind, so that the way into it has, I think, been made quite easy.

A novel feature of the latter part of the volume is the discussion of the remarkable analogy between geometrical optics and Hamiltonian dynamics and the use made of it, by expanding it so that it becomes an analogy between optics (using this term in its widest sense and as understood before the days of the quantum theory) and dynamics, to develop quantum dynamics.

In writing the volume I have of course made much use of the works of authoritative writers on the subjects dealt with : in relativity, of the original papers of Einstein and Minkowski and of Eddington's works ; in the quantum theory, of the papers of the great master Planck as well as of those of Bohr, Sommerfeld, de Broglie, Heisenberg, and Schroedinger, and of Dirac's Quantum Mechanics.

The volume is practically self-contained, though there are naturally many references to the earlier volumes, and the numbering of the sections follows on from that of the preceding volume.

I wish to express my indebtedness to Dr. R. C. Johnson for the figure illustrating the Stark effect and to Dr. Jessie Cattermole for the trouble she has taken in helping me with the proofs.

W. W.

October 1939

THEORETICAL PHYSICS, III

CHAPTER I

THEORY OF RELATIVITY

§ 34. The 4-Dimensional Character of Field Equations

WE have noticed occasionally that the equations describing vector or tensor fields can be given a 4-dimensional form. The sets of equations **(10·55)**, **(22·74)**, and **(27·47)** are instructive examples of this. The first of them describes the force per unit volume associated with stress and momentum in an elastic material medium, the second set consists of the electromagnetic field equations, while the third one determines the electromagnetic vector and scalar potentials. We should now give closer attention to this 4-dimensional character, more particularly to the manifestation of it which appears in the electromagnetic field equations. These may be written in the form :

$$\left.\begin{aligned} \frac{4\pi A}{a}\frac{\partial \mathbf{D}}{\partial t} + \frac{4\pi A\rho \mathbf{v}}{a} &= \mathbf{curl}\ \mathbf{H}, \\ \mathbf{div}\ \mathbf{D} &= \rho, \\ -\frac{4\pi A}{a}\frac{\partial \mathbf{D}_m}{\partial t} &= \mathbf{curl}\ \mathfrak{E}, \\ \mathbf{div}\ \mathbf{D}_m &= 0. \end{aligned}\right\} \quad . \ . \ . \quad (34)$$

In this form they are suitable for a region filled with a homogeneous isotropic insulating medium containing charges (charged particles) which can move freely through it. The simplest insulating medium is empty space—if we may be permitted to apply the term ' medium ' to empty space—and by this is meant a region which has been exhausted by a pump, or which is empty in the same sense as such a region. We shall leave the question open for the present as to whether so-called empty space is occupied by some plenum which the exhausting devices of experimental physicists fail to grasp, and apply to it equations **(34)**. The symbols $\mathbf{D}$ and $\mathbf{D}_m$ signify the electric and magnetic displacement respectively at some point ; ρ is the charge density at this point and $\mathbf{v}$ its velocity of convection, while $\mathfrak{E}$ and $\mathbf{H}$ are respectively the electric and magnetic field intensities. Finally A and a are the purely numerical quantities already

described in §§ **19·1** and **22·1**. Our choice of units usually gives A one or other of the values 1 or $1/4\pi$, while a is usually either unity or has the same numerical value as the velocity, c, of electromagnetic waves in empty space.

Each of the first and third of the equations **(34)** is really an abbreviated statement of three equations, so that **(34)** contains two sets of four equations. The first of these sets may be written in the form (cf. **22·74**):

$$\left.\begin{aligned}
\frac{\partial F^{11}}{\partial x^1}+\frac{\partial F^{12}}{\partial x^2}+\frac{\partial F^{13}}{\partial x^3}+\frac{\partial F^{14}}{\partial x^4}&=s^1,\\
\frac{\partial F^{21}}{\partial x^1}+\frac{\partial F^{22}}{\partial x^2}+\frac{\partial F^{23}}{\partial x^3}+\frac{\partial F^{24}}{\partial x^4}&=s^2,\\
\frac{\partial F^{31}}{\partial x^1}+\frac{\partial F^{32}}{\partial x^2}+\frac{\partial F^{33}}{\partial x^3}+\frac{\partial F^{34}}{\partial x^4}&=s^3,\\
\frac{\partial F^{41}}{\partial x^1}+\frac{\partial F^{42}}{\partial x^2}+\frac{\partial F^{43}}{\partial x^3}+\frac{\partial F^{44}}{\partial x^4}&=s^4.
\end{aligned}\right\} \quad . \quad . \quad \textbf{(34·01)}$$

The co-ordinate symbols x, y, and z have been replaced by x^1, x^2, and x^3 respectively, while x^4 represents a certain product, c_0t, where c_0 is a constant velocity, which for the present may remain undefined; but to which we shall be able to assign a meaning and a value later. Its sole function meanwhile is to enable us to achieve the symmetry shown in equations **(34·01)**. The symbols, $F^{\alpha\beta}$, have the property:

$$F^{\alpha\beta} = -F^{\beta\alpha},$$

so that

$$F^{\alpha\alpha} = 0,$$

and

$$\left.\begin{aligned}
&F^{12}=H_z,\ F^{31}=H_y,\ F^{23}=H_x,\\
&F^{41}=\frac{4\pi A c_0 D_x}{a},\ F^{42}=\frac{4\pi A c_0 D_y}{a},\ F^{43}=\frac{4\pi A c_0 D_z}{a},\\
&s^1=\frac{4\pi A\rho v_x}{a},\ s^2=\frac{4\pi A\rho v_y}{a},\ s^3=\frac{4\pi A\rho v_z}{a},\\
&s^4=\frac{4\pi A c_0\rho}{a}.
\end{aligned}\right\} \quad \textbf{(34·011)}$$

The substitution of these expressions for $\mathbf{D}$, $\mathbf{D}_m$, ρ, and $\mathbf{v}$, or for their components, in **(34)** will be seen to yield the equations **(34·01)**.

The summation convention, often employed already, enables

us to abbreviate very considerably the statement of equations **(34·01)**. The third equation, for example, may be written:

$$\frac{\partial F^{3\alpha}}{\partial x^{\alpha}} = s^{3},$$

and any one of the four equations may be expressed in the form:

$$\frac{\partial F^{\beta\alpha}}{\partial x^{\alpha}} = s^{\beta}. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(34·012)}$$

We may write the third and fourth of the equations **(34)** in the following way:

$$\left.\begin{aligned} \frac{\partial F_{43}}{\partial x^{2}} + \frac{\partial F_{24}}{\partial x^{3}} + \frac{\partial F_{32}}{\partial x^{4}} &= 0, \\ \frac{\partial F_{41}}{\partial x^{3}} + \frac{\partial F_{34}}{\partial x^{1}} + \frac{\partial F_{13}}{\partial x^{4}} &= 0, \\ \frac{\partial F_{42}}{\partial x^{1}} + \frac{\partial F_{14}}{\partial x^{2}} + \frac{\partial F_{21}}{\partial x^{4}} &= 0, \\ \frac{\partial F_{13}}{\partial x^{2}} + \frac{\partial F_{21}}{\partial x^{3}} + \frac{\partial F_{32}}{\partial x^{1}} &= 0, \end{aligned}\right\} \quad . \quad . \quad . \quad \textbf{(34·02)}$$

the symbols having the meanings:

$$\left.\begin{aligned} F_{32} = \varepsilon\frac{4\pi A c_0 D_{mx}}{a}, \quad F_{13} = \varepsilon\frac{4\pi A c_0 D_{my}}{a}, \quad F_{21} = \varepsilon\frac{4\pi A c_0 D_{mz}}{a}, \\ F_{41} = \varepsilon\mathcal{E}_x, \quad F_{42} = \varepsilon\mathcal{E}_y, \quad F_{43} = \varepsilon\mathcal{E}_z. \end{aligned}\right\} \quad . \quad \textbf{(34·021)}$$

For the present ε is any constant. We shall deal with its value and dimensions later. We have assigned the numerical subscripts to the components F in **(34·02)** on the principle that each component, for example F^{41}, in **(34·01)** is proportional (when μ and K are constants) to the corresponding component, for example F_{41}, in **(34·02)**. Corresponding components like F^{41} and F_{41} are distinguished from one another by the positions of the indices 4, 1.

In the case of empty space μ and K are constants which we may represent by μ_0 and K_0 respectively, and since

$$D_{mx} = \frac{\mu_0 H_x}{4\pi A},$$

it follows that, in empty space (cf. **34·011** and **34·021**),

$$F_{32} = -\varepsilon\frac{c_0\mu_0}{a}F^{32}, \quad . \quad . \quad . \quad . \quad \textbf{(34·03)}$$

and there are similar expressions for F_{21} and F_{13}. Further,

$$F_{41} = \varepsilon \frac{a}{c_0 K_0} F^{41}, \quad . \quad . \quad . \quad . \quad . \quad (34{\cdot}04)$$

in empty space.

The suggestion now arises that F^{23}, F^{31}, etc., as well as F_{23}, F_{31}, etc., constitute tensors of the second rank in a 4-dimensional continuum and we shall try the experiment of regarding them as the same tensor, in empty space, apart from a constant factor representing the ratio of a component of one of them, say F^{32}, to the corresponding component, F_{32}, of the other. The constant factor is of course

$$-\varepsilon \frac{c_0 \mu_0}{a} = \varepsilon \frac{a}{c_0 K_0}. \quad . \quad . \quad . \quad . \quad . \quad (34{\cdot}05)$$

We thus obtain

$$c_0{}^2 = -\frac{a^2}{\mu_0 K_0},$$

or

$$c_0 = i \frac{a}{\sqrt{\mu_0 K_0}},$$

or finally

$$c_0 = ic, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (34{\cdot}06)$$

where c is the velocity of electromagnetic waves in empty space and $i = \sqrt{-1}$.

The constant, ε, is at our disposal and it is convenient to assign to it the value:

$$\varepsilon = i \frac{a}{4\pi A \,.\, c} \quad . \quad . \quad . \quad . \quad . \quad (34{\cdot}061)$$

which has the dimensions of a reciprocal velocity. On introducing this value in **(34·021)** we find that

$$\left.\begin{aligned} F_{23} &= D_{mx}, \\ F_{31} &= D_{my}, \\ F_{12} &= D_{mz}, \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (34{\cdot}07)$$

while

$$\left.\begin{aligned} F_{41} &= \frac{ia}{4\pi A c} \mathfrak{E}_x, \\ F_{42} &= \frac{ia}{4\pi A c} \mathfrak{E}_y, \\ F_{43} &= \frac{ia}{4\pi A c} \mathfrak{E}_z. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (34{\cdot}071)$$

The apparently arbitrary assignment of the value **(34·061)** to ε will be justified in **§ 35·2.**

From **(34·011)** we have

$$\left.\begin{aligned} F^{23} &= H_x, \\ F^{31} &= H_y, \\ F^{12} &= H_z, \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad \textbf{(34·08)}$$

while the substitution of ic for c_0 yields

$$\left.\begin{aligned} F^{41} &= i\frac{4\pi Ac}{a}D_x, \\ F^{42} &= i\frac{4\pi Ac}{a}D_y, \\ F^{43} &= i\frac{4\pi Ac}{a}D_z. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad \textbf{(34·081)}$$

We have decided that we shall regard $F^{\alpha\beta}$ and $F_{\alpha\beta}$ as the components of tensors of rank 2 in a 4-dimensional continuum.

The expressions for s^1, s^2, s^3, and s^4 in **(34·01)** have the familiar divergence form which we have met (**§ 9·9**) in the equations for the force per unit volume due to stress and elsewhere, and we shall therefore regard them as the components of a vector in a 4-dimensional continuum of a Euclidean character.

§ 34·1. Galilean Space-Time

In Euclidean space a change from one system of rectangular co-ordinates, X, Y, Z, to another, X', Y', Z', leaves the square of the line element, $dx^2 + dy^2 + dz^2$, invariant, i.e.

$$dx^2 + dy^2 + dz^2 = dx'^2 + dy'^2 + dz'^2,$$

and indeed

$$x^2 + y^2 + z^2 = x'^2 + y'^2 + z'^2,$$

when the new axes, X', Y', Z', are simply the result of turning the old ones about the origin and (x, y, z) and (x', y', z') are the co-ordinates of the same point in the respective co-ordinate systems.

The results of the last section suggest a kind of fusion of space and time [1] into a single 4-dimensional continuum—we shall call it **Galilean space-time**—in which the square of the line element is

$$(dx^1)^2 + (dx^2)^2 + (dx^3)^2 + (dx^4)^2 \quad . \quad . \quad \textbf{(34·1)}$$

Just as in Euclidean space the square of the line element, $(dx^1)^2 + (dx^2)^2 + (dx^3)^2$ or, as we more usually write it, $dx^2 + dy^2 + dz^2$ is invariant under co-ordinate transformations,

[1] Cf. H. Minkowski : *Raum und Zeit.*

so we shall now lay down that the expression **(34·1)** is an invariant. The typical vector in the new continuum—we may call it a displacement—has the components:

$$dx^1,\ dx^2,\ dx^3,\ dx^4,$$

where dx^1, dx^2, and dx^3 are the components of the displacement referred to rectangular axes of co-ordinates $(X,\ Y,\ Z)$ in space, while

$$dx^4 = c_0\, dt = ic\, dt, \quad . \ . \ . \ . \ \textbf{(34·101)}$$

where dt means the time during which the spacial part

$$(dx^1,\ dx^2,\ dx^3) \text{ or } (dx,\ dy,\ dz)$$

of the displacement occurs.

Since $(dx^4)^2$ is negative, it is possible for the square of the line element to vanish even when its individual components do not all separately vanish; in which case

$$(dx^1)^2 + (dx^2)^2 + (dx^3)^2 = c^2(dt)^2,$$

or

$$\frac{(dx^1)^2 + (dx^2)^2 + (dx^3)^2}{(dt)^2} = c^2 \quad . \ . \ . \ \textbf{(34·11)}$$

This represents the special case where the spacial displacement is occurring with the velocity, c, of electromagnetic waves (light) in empty space. The invariance of the expression **(34·1)** necessitates that, if it vanishes in one system of reference, it must vanish in all, and consequently **when anything whatever is travelling with this particular velocity, c, its velocity will be equal to c in all cases, whatever may be the frame of reference relative to which it is measured.**

We shall distinguish the axes of co-ordinates in the 4-dimensional continuum by $(X^1,\ X^2,\ X^3,\ X^4)$, or sometimes, when this seems desirable, by $(X,\ Y,\ Z,\ W)$, and when we wish to distinguish the new mode of representing physical phenomena from the older one, we shall mark the former by the letter G and the latter by E. A fixed point in the E representation will obviously be, in the G representation, a straight line parallel to the X^4 (or W) axis,[1] since its spacial co-ordinates $(x,\ y,\ z)$ remain the same for all values of t (or w) and therefore it is the locus represented by particular values of x^1, x^2, and x^3 and all values of x^4. A rotation of the G axes about the origin may cause this straight line to be inclined to the X^4 (or W) axis, so that in the G representation the spacial co-ordinates of the points on it increase (or decrease) with increasing values of x^4 (or w), and

[1] The precise meanings of ' straight line ', ' parallel ', ' rotation ', etc., in Galilean space-time are defined below.

at any point on it—this means of course a spacial point at a particular instant—dx^1/dx^4 or $dx/dw = dx/ic\,dt = v_x/ic$ and is constant. In the old-fashioned E representation, therefore, we now have, not a point at rest, but one moving with a constant velocity (v_x, v_y, v_z). **Such a rotation of the G axes about the origin is equivalent, therefore, in the E representation, to a change from one set of rectangular axes of co-ordinates to another moving relatively to it with a constant velocity of translation.**

The terms ' straight line ', ' parallel ', ' rotation ', etc., have hitherto been used exclusively in regard to things in Euclidean space in which they have a certain visual significance which is absent when we apply them to relationships involving the X^4 (or W) axis in the Galilean continuum. These terms are used in connexion with the 4-dimensional continuum for things whose mathematical description is identical with that of the things so denominated in Euclidean space.

We have already noticed that the particular velocity, c, namely that of light, or of electromagnetic waves, in empty space, is not modified by such co-ordinate changes as we are now contemplating, and we have thus at the very outset found a way of fitting the negative result of Michelson's famous experiment **(§ 33·4)** into a place in a new and wider scheme.

§ 34·2. The Lorentz Transformation

A rotation, if we may call it such, of a Euclidean set of rectangular axes of co-ordinates about the origin, which leaves the Y and Z co-ordinates of a given point unaltered, i.e. one which makes

$$y' = y$$

and

$$z' = z,$$

must leave

$$x'^2 = x^2,$$

and consequently

$$x' = +x \text{ or } -x.$$

We have already met with these two cases in **§ 2·1** and we there learned that the change from x to $-x$ is really a change from one kind of set of rectangular axes to another and not a rotation at all. Similarly when in the G representation we leave the co-ordinates y, z, and w of a point-instant unchanged it will necessitate that its X co-ordinate remains unchanged.

Let us next consider a G rotation about the origin, in which

the co-ordinates z and w of a point-instant, P (Fig. 34·2), remain unchanged. Obviously in this case $x'^2 + y'^2 = x^2 + y^2$, and the change is one which, in the E representation, would be called a rotation about the Z axis. The equations of transformation are in such a case :

$$\left.\begin{aligned} x' &= x\cos\varepsilon - y\sin\varepsilon, \\ y' &= x\sin\varepsilon + y\cos\varepsilon \\ z' &= z, \\ w' &= w, \end{aligned}\right\} \quad . \quad . \quad . \quad (34\cdot2)$$

the angle ε being that shown in the figure.

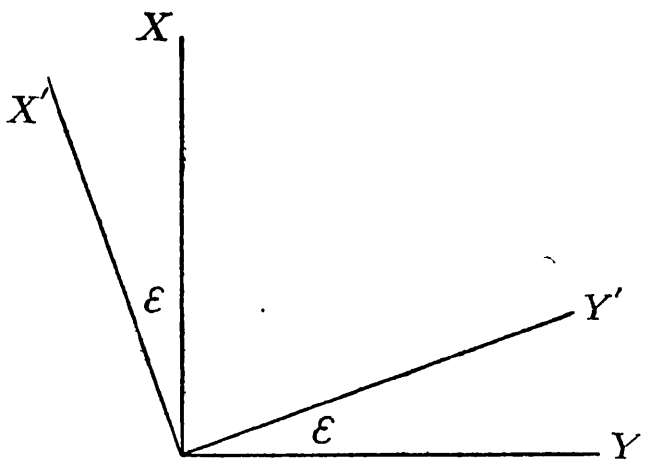

FIG. 34·2

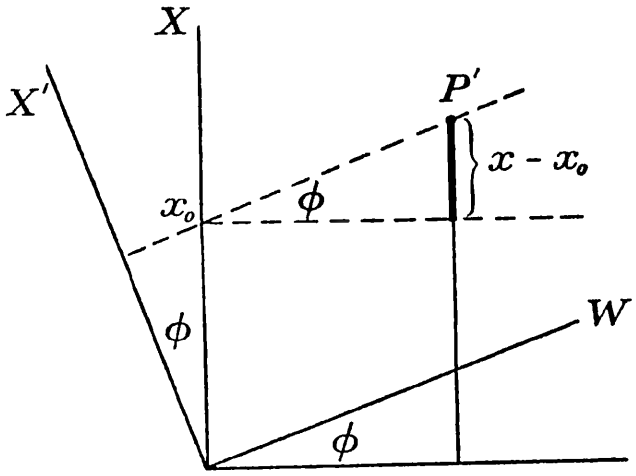

FIG. 34·21

We now proceed to study the case of a rotation of the G axes, about the origin, where two of the spacial co-ordinates of a point-instant are kept constant—say y and z—so that

$$\left.\begin{aligned} y' &= y, \\ z' &= z. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (34\cdot21)$$

and

We can deal with this case very shortly and easily. Let us imagine, in Fig. 34·2, Y and Y' replaced by W and W' respectively and the angle ε by ϕ, as in Fig. 34·21. We thus learn that

$$\left.\begin{aligned} x' &= x\cos\phi - w\sin\phi, \\ w' &= x\sin\phi + w\cos\phi. \end{aligned}\right\} \quad . \quad . \quad . \quad (34\cdot211)$$

Of course the figure as now used has no other significance for us than that of shortly indicating the equations of transformation **(34·211)**. Indeed, ϕ is not a real angle. Imagine a straight line x_0P' passing through the projection, P', of P on the XW or $X'W'$ plane and parallel to the W' axis. Let it cut the X axis at $x = x_0$. And imagine another straight line through x_0 and parallel to the W axis. It will be seen that the angle between these two lines is ϕ and that

$$\tan\phi = (x - x_0)/w = (x - x_0)/ict \quad . \quad . \quad (34\cdot212)$$

Now the points (point-instants) on the line x_0P' are points whose X' co-ordinate retains the same value as w' increases. They represent a spacial point *at rest* in the (X', Y', Z') Euclidean co-ordinates. Hence

$$(x - x_0)/t$$

represents the (constant) velocity, v, with which the new Euclidean axes (X', Y', Z') are moving relatively to the old ones (X, Y, Z). So we may write instead of **(34·212)**

$$\tan \phi = v/ic. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(34·213)}$$

It follows at once from this equation that

$$\cos \phi = (1 - v^2/c^2)^{-\frac{1}{2}} \quad . \quad . \quad . \quad . \quad \textbf{(34·214)}$$

Thus $\cos \phi$ is identical with the function of v (cf. **§ 27·2**) we have already denoted by γ. Clearly

$$\sin \phi = v \cos \phi/ic$$

or

$$\sin \phi = \gamma v/ic \quad . \quad . \quad . \quad . \quad . \quad \textbf{(34·215)}$$

On substituting these expressions for $\cos \phi$ and $\sin \phi$ in **(34·211)**, and ict and ict' for w and w' respectively, we get, if we include equations **(34·21)**,

$$\left.\begin{aligned} x' &= \gamma(x - vt), \\ y' &= y, \\ z' &= z, \\ t' &= \gamma(t - vx/c^2). \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad \textbf{(34·22)}$$

These equations constitute the **Lorentz transformation,** so named in honour of H. A. Lorentz.[1] The transformation was also discovered by W. Voigt [2] as early as 1887. Certain considerations described in **§ 28** suggested the first of these equations, but the fourth one is new. They arise quite naturally out of the adoption of the 4-dimensional Galilean continuum.

The inverse transformation corresponding to **(34·21)** **(34·211)** is obviously expressed by:

$$\left.\begin{aligned} x &= x' \cos \phi + w' \sin \phi, \\ y &= y', \\ z &= z', \\ w &= x' (- \sin \phi) + w' \cos \phi. \end{aligned}\right\} \quad . \quad . \quad . \quad \textbf{(34·23)}$$

[1] H. A. Lorentz : *Proc. Acad. Sc. Amsterdam,* 6 (1904), p. 809. Also included in *Das Relativitätsprinzip.*

[2] W. Voigt : *Göttinger Nachriehungen,* p. 41 (1887).

That corresponding to **(34·22)** is

$$\left.\begin{aligned} x &= \gamma(x' + vt'), \\ y &= y', \\ z &= z', \\ t &= \gamma\left(t' + \frac{vx'}{c^2}\right). \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad \textbf{(34·24)}$$

In the E representation, **(34·22)** expresses the transformation from rectangular co-ordinates (X, Y, Z) to rectangular co-ordinates (X', Y', Z'), which are moving with a constant velocity, v, relatively to the former in the direction of X (or X'), the two sets of co-ordinates being coincident at the instant $t = t' = 0$.

In the old-fashioned Newtonian physics we have, of course, instead of **(34·22)**, the equations of transformation:

$$\left.\begin{aligned} x' &= x - vt, \\ y' &= y, \\ z' &= z, \\ t' &= t. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad \textbf{(34·25)}$$

These equations are practically identical with **(34·22) when v is very small compared with c,** as indeed we should expect.

§ 34·3. Some Applications of the Lorentz Transformation

Imagine a particle moving in the X, Y, Z (E) system with the velocity (u_x, u_y, u_z) and in the X', Y', Z' system with the velocity (u_x', u_y', u_z'). We inquire about the relationship between these two velocities. From **(34·22)** we have

$$u_x' = \frac{dx'}{dt'} = \frac{\gamma(dx - v\,dt)}{\gamma\left(dt - \frac{v\,dx}{c^2}\right)},$$

$$u_y' = \frac{dy'}{dt'} = \frac{dy}{\gamma\left(dt - \frac{v\,dx}{c^2}\right)},$$

$$u_z' = \frac{dz'}{dt'} = \frac{dz}{\gamma\left(dt - \frac{v\,dx}{c^2}\right)};$$

or

$$\left.\begin{aligned} u_x' &= \frac{u_x - v}{1 - \frac{vu_x}{c^2}}, \\ u_y' &= \frac{u_y}{\gamma\left(1 - \frac{vu_x}{c^2}\right)}, \\ u_z' &= \frac{u_z}{\gamma\left(1 - \frac{vu_x}{c^2}\right)}, \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (34\cdot3)$$

instead of the familiar

$$\begin{aligned} u_x' &= u_x - v, \\ u_y' &= u_y, \\ u_z' &= u_z. \end{aligned}$$

These kinematical equations **(34·3)** were first given by Einstein.[1] We shall now make some applications of them and of the Lorentz transformation from which they emerge. Imagine a rod fixed in the X', Y', Z' (E) co-ordinate system and parallel to the X' axis. The co-ordinates of its two ends might be, for example, $(x_1', 0, 0)$ and $(x_2', 0, 0)$. Its length referred to this co-ordinate system, assuming $x_2' > x_1'$, is therefore

$$L' = x_2' - x_1'.$$

On the other hand, its length referred to the X, Y, Z (E) system of co-ordinates is $x_2 - x_1$, where x_2 and x_1 are the X co-ordinates of its ends, *both corresponding to the same instant, t.* It follows from **(34·22)** that

$$x_2' - x_1' = \gamma(x_2 - x_1)$$

or

$$L' = \gamma L, \quad . \quad . \quad . \quad . \quad . \quad . \quad (34\cdot31)$$

where L is the length of the rod referred to the X, Y, Z (E) system, relatively to which it is moving with the velocity, v. The relation **(34·31)** has already appeared, rather tentatively, in **§§ 28** and **33·4** as the **FitzGerald-Lorentz contraction hypothesis.** It emerges quite naturally from our new 4-dimensional way of representing physical phenomena. The term ' contraction ', we now realize, is inappropriate ; the fact being that the dimensions of a body, just as much as its velocity, depend

[1] A. Einstein : *Ann. d. Physik,* 17 (1905), p. 891. Included in *Das Relativitätsprinzip.*

on the system of co-ordinates in which they are measured or expressed.

It is instructive to derive **(34·31)** directly from the G representation, instead of using the Lorentz transformation. The X, W and X', W' axes of co-ordinates are shown in Fig. 34·3. The rod in question is fixed in the X', Y', Z' (E) system. This obviously means that its two ends are on two lines parallel to the W' axis (Fig. 34·3), since the values of x_1' and x_2' are constants independent of w'. Obviously L and L' are the lengths illustrated in the figure, since L means the difference, $x_2 - x_1$, both x_1 and x_2 corresponding to the same instant, t, i.e. to the same value of w. The figure shows that

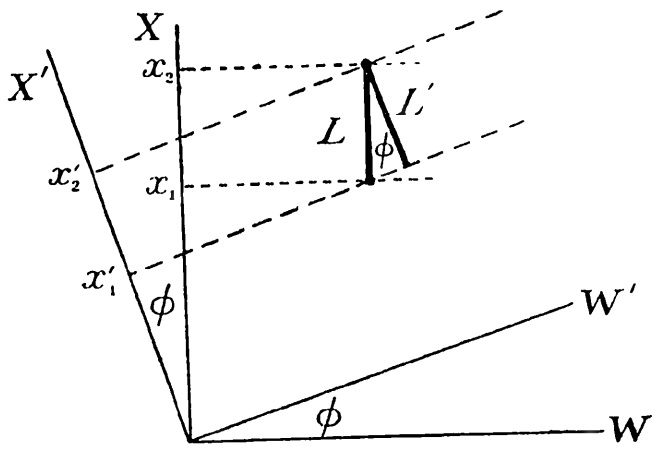

FIG. 34·3

$$L' = L\cos\phi,$$

or

$$L' = \gamma L,$$

which is the result we have already found.

Consider next a clock fixed in the X', Y', Z' (E) system of co-ordinates and used to measure the interval of time,

$$T' = t_2' - t_1',$$

between two events. We can get the time interval between these two events, referred to the X, Y, Z (E) co-ordinates, from the last of the equations **(34·24)**. We have

$$t_1 = \gamma\left(t_1' + \frac{vx'}{c^2}\right),$$

$$t_2 = \gamma\left(t_2' + \frac{vx'}{c^2}\right),$$

since the clock, being at rest in the X', Y', Z' (E) system, has the same X' co-ordinate, x', at both instants t_1' and t_2'. Hence

$$T = t_2 - t_1 = \gamma(t_2' - t_1'),$$

or

$$T = \gamma T'.$$

We can get this result directly from the G representation with the aid of Fig. 34·31. The position of the clock is represented by the line through x' parallel to the W' axis. The points

(point-instants) w_1' and w_2' indicate the terminal instants of $T' = t_2' - t_1'$, while w_1 and w_2 indicate the terminal instants of $T = t_2 - t_1$. The figure shows at once that

$$w_2 - w_1 = \cos\phi(w_2' - w_1'),$$

and therefore

$$ic(t_2 - t_1) = ic\gamma(t_2' - t_1'),$$

or

$$T = \gamma T'$$

as we have already found. This result asserts that a clock which is in motion relatively to an observer will give for the time interval between two events a smaller quantity than will the same clock when fixed relatively to him. It will appear, to observers in X, Y, Z, to go too slowly. This is the time analogue of the FitzGerald-Lorentz contraction. The truth is that both length and time have now lost their absolute significance, having passed it on to a sort of combination of length and time, namely, the interval between two point-instants in the 4-dimensional Galilean continuum, the square of which is expressed by **(34·1)**.

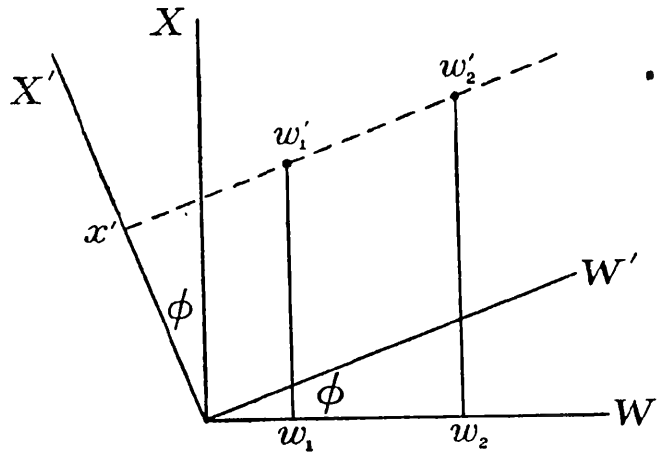

FIG. 34·31

§ 34·4. THE ABERRATION AND CONVECTION OF LIGHT—DOPPLER EFFECT

The negative result of Michelson's experiment **(§§ 31·3** and **33·4)** is an essential part of the very foundation of our new theory **(§ 34·1)** and may now be dismissed. We turn next to certain other important phenomena associated with the propagation of light—more especially the aberration and convection of light, which before the advent of the theory of relativity could not be satisfactorily accounted for, nor entirely reconciled with Michelson's result.

Imagine a homogeneous and isotropic medium, fixed relatively to the X, Y, Z (E) axes and a plane harmonic light wave, or electromagnetic wave, travelling in the direction (l, m, n).

The equation describing the wave may be written in the form:

$$s = A \cos 2\pi\nu\left(t - \frac{lx + my + nz}{c/\mu}\right), \quad . \quad . \quad \textbf{(34·4)}$$

in which the meanings of the symbols need no elucidation. It should be emphasized, however, that everything in the equation, including the refractive index, μ, is referred to the X, Y, Z system of reference in which the medium is, by hypothesis, at rest. Let us inquire how the wave **(34·4)** appears when referred to a system X', Y', Z' (E) coincident with X, Y, Z at the instant $t = t' = 0$ and moving relatively to the latter system with the velocity, v, in the common X or X' direction. We may write at once

$$s = A \cos 2\pi\nu\left\{\gamma\left(t' + \frac{vx'}{c^2}\right) - \frac{l\gamma(x' + vt') + my' + nz'}{c/\mu}\right\},$$

if we make use of **(34·24)**; and it should be remembered that s, A and μ, the values of the displacement, amplitude and refractive index respectively, are still referred to the original axes. Therefore

$$s = A \cos 2\pi\nu'\{t' - (L'x' + M'y' + N'z')\}, \quad . \quad \textbf{(34·41)}$$

where

$$\left.\begin{aligned} \nu' &= \gamma\left(1 - \frac{lv\mu}{c}\right)\nu, \\ L' &= \frac{\mu\left(l - \frac{v}{c\mu}\right)}{c\left(1 - \frac{lv\mu}{c}\right)}, \\ M' &= \frac{\mu m}{\gamma c\left(1 - \frac{lv\mu}{c}\right)}, \\ N' &= \frac{\mu n}{\gamma c\left(1 - \frac{lv\mu}{c}\right)}. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad \textbf{(34·42)}$$

The first of these four equations expresses the Doppler change in frequency

$$-\delta\nu = \nu - \nu',$$

which, if we neglect second order small quantities—v/c is a small

quantity—and suppose that l is not a second order small quantity, becomes

$$-\delta\nu = \frac{l\nu\mu}{c}\nu,$$

or, since $\nu = 1/\tau$,

$$\delta\tau = -\delta\nu/\nu^2,$$

and therefore

$$\delta\tau = \frac{l\nu\mu}{c}\tau, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(34·43)}$$

and when the wave is travelling in the common X or X' direction, i.e. when $l = 1$,

$$d\tau = \nu\mu\tau/c, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(34·431)}$$

which is the expression used in **§ 33·2.**

If c' be the phase velocity of the wave in the X', Y', Z' system, then

$$1/c'^2 = L'^2 + M'^2 + N'^2,$$

and when the wave is travelling in the common X or X' direction ($l = 1$ and $m = n = 0$)

$$\frac{1}{c'^2} = L'^2 = \frac{\mu^2\left(1 - \frac{v}{c\mu}\right)^2}{c^2\left(1 - \frac{v\mu}{c}\right)^2},$$

and therefore

$$c' = \frac{c(1 - v\mu/c)}{\mu(1 - v/c\mu)}$$

Neglecting second order small quantities, this is equivalent to

$$c' = \frac{c}{\mu}\left(1 - \frac{v\mu}{c} + \frac{v}{c\mu}\right),$$

or

$$c' = \frac{c}{\mu} - v\left(1 - \frac{1}{\mu^2}\right) \quad . \quad . \quad . \quad . \quad \textbf{(34·44)}$$

This is the convection formula **(33·26)** discovered by Fresnel.

It will be remembered that μ is the refractive index as referred to the system X, Y, Z. If μ' be the refractive index referred to X', Y', Z',

$$\mu' = \mu + \frac{d\mu}{d\tau}d\tau,$$

or by **(34·431)**,

$$\mu' = \mu + \frac{d\mu}{d\tau}\cdot\frac{v\mu\tau}{c}.$$

The second term is small and, since we are neglecting second order small quantities, we may write the equation as

$$\mu = \mu' - \frac{v\mu'\tau'}{c}\frac{d\mu'}{d\tau}.$$

We substitute this for μ in the first term only (since we are neglecting second order small quantities) of **(34·44)** and replace μ in the small term by μ'. Thus

$$c' = \frac{c}{\mu' - \dfrac{\mu' v\tau'}{c}\dfrac{d\mu'}{d\tau'}} - v\left(1 - \frac{1}{\mu'^2}\right)$$

or

$$c' = \frac{c}{\mu'} - v\left(1 - \frac{1}{\mu'^2} - \frac{\tau'}{\mu'}\frac{d\mu'}{d\tau'}\right) \quad . \quad . \quad \textbf{(34·441)}$$

The formulae **(34·44)** and **(34·441)** have already been given in **§ 33·2.** Their derivation was there based on the observational fact that the laws of refraction are independent of the relative motion of the medium and the source of light, to the first order of approximation. We have in the present derivation the great advantage that we have made no assumptions which have a peculiar application to the propagation of light; but only such as are inherent in the foundation of our new theory and which apply equally to all physical phenomena.

When the wave **(34·4)** is travelling in the Y direction ($m = 1$, $l = n = 0$) the equations **(34·42)** become

$$\left.\begin{aligned} \nu' &= \gamma\nu, \\ L' &= -v/c^2, \\ M' &= \mu/\gamma c, \\ N' &= 0, \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad \textbf{(34·45)}$$

and the direction cosines of the direction of propagation relatively to the X', Y', Z' system are

$$\left.\begin{aligned} l' &= \frac{-\dfrac{v}{c^2}}{\sqrt{\dfrac{v^2}{c^4} + \dfrac{\mu^2}{\gamma^2 c^2}}}, \\ m' &= \frac{\dfrac{\mu}{\gamma c}}{\sqrt{\dfrac{v^2}{c^4} + \dfrac{\mu^2}{\gamma^2 c^2}}}, \\ n' &= 0, \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad \textbf{(34·46)}$$

These equations express the aberration of light. For inter-stellar space, in particular, $\mu = 1$, and they become

$$\left.\begin{aligned} l' &= -v/c, \\ m' &= 1/\gamma, \\ n' &= 0. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (34{\cdot}47)$$

When the light is travelling from a star fixed in the X, Y, Z system and situated at $x = 0$, $y = -\infty$, $z = 0$, the light waves will be plane waves at the origin, O, and although not mono-

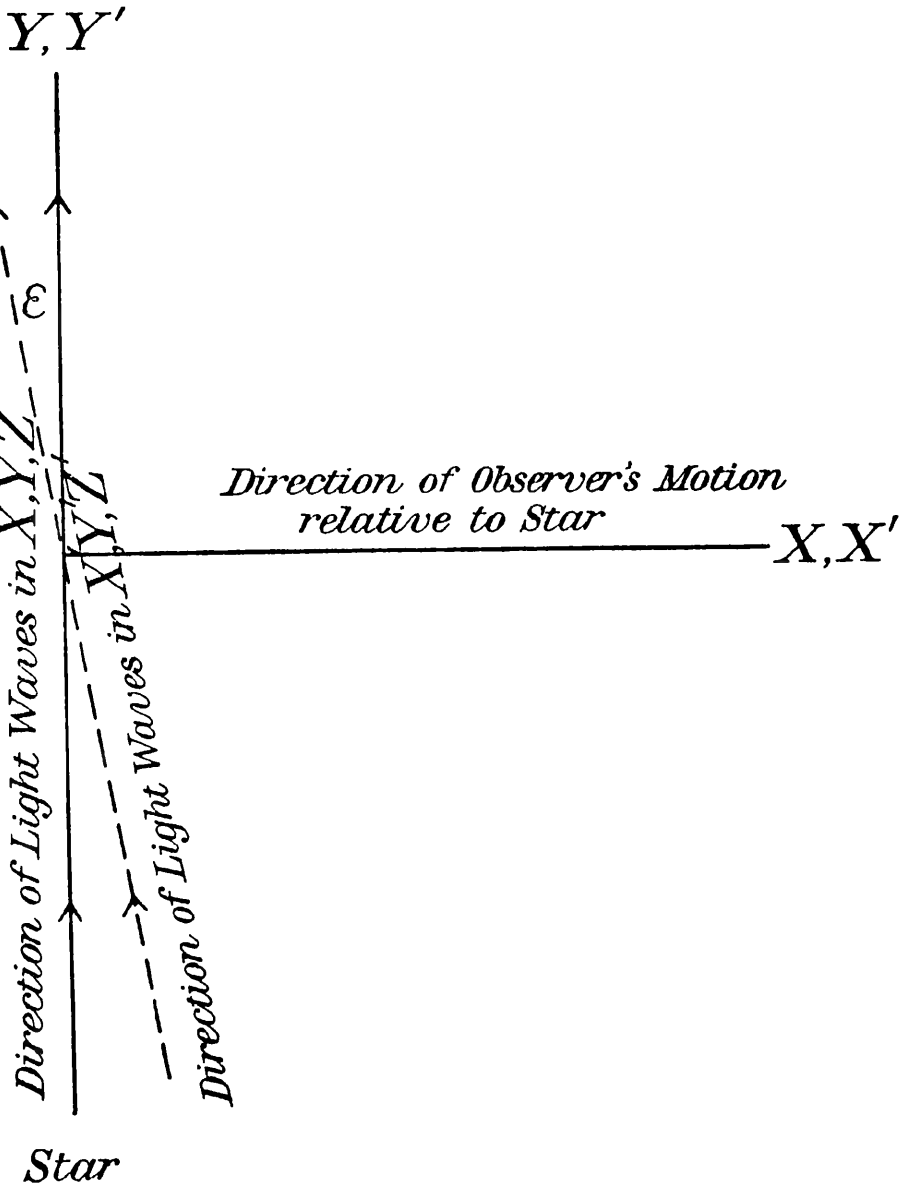

FIG. 34·4

chromatic waves they may be represented as a superposition of a large (or infinite) number of terms like **(34·4)** in all of which $\mu = 1$, and it follows that the direction cosines of the composite wave, in the X', Y', Z' system in which the observer on the earth may be supposed to be fixed, will be those described by **(34·47)**. He will find the direction of the wave to be perpen-

dicular to the Z' axis ($n' = 0$) and to make the small angle, ε, (Fig. 34·4) with the Y' axis as shown in the figure; so that, since

$$l' = -v/c,$$
$$\sin \varepsilon = v/c, \quad . \quad . \quad . \quad . \quad . \quad (34{\cdot}48)$$

and we have an expression for stellar aberration, which is in agreement with the observations and entirely consistent with the negative result of Michelson's experiment (cf. **§ 33·3**).

§ 34·5. The Principle of Relativity

At this point we shall look back and review the ground we have covered. Acting on the hint given by the 4-dimensional form of many important sets of equations—more especially Clerk-Maxwell's electromagnetic field equations—we have been led to treat space and time as a single 4-dimensional continuum, and not as two quite separate and disconnected continua. This new 4-dimensional continuum, which we have called a **Galilean continuum**, may be said to have the mathematical characteristics of the Euclidean continuum extended to a further dimension. Apart from its 4-dimensional character it also differs from a Euclidean continuum in a certain particular which, though mathematically trivial, is of cardinal importance from a physical point of view: this difference lies in the fact that the square of the interval **(34·1)** can vanish even when its constituent terms do not separately do so. It is this peculiarity which has enabled us to solve the problems associated with Michelson's experiment.

In pre-relativistic physics the equations describing physical phenomena are equations connecting Euclidean vector or tensor quantities, and these equations have precisely the same form in all systems of reference which are fixed relatively to one another—fixed, we might say, in the same Euclidean space. For instance, the expression **(9·91)**, namely,

$$\frac{\partial t_{xx}}{\partial x} + \frac{\partial t_{xy}}{\partial y} + \frac{\partial t_{xz}}{\partial z}, \quad . \quad . \quad . \quad . \quad (34{\cdot}5)$$

which represents the X component of the force per unit volume due to a condition of stress in a homogeneous isotropic elastic solid, referred to rectangular axes of co-ordinates X, Y, Z, will also represent the same component of the force per unit volume in any other set of rectangular axes of co-ordinates X', Y', Z' (fixed relatively to X, Y, Z), provided we replace t_{xx}, t_{xy}, etc., by the corresponding components t_{xx}', t_{xy}', etc., of the same

tensor referred to the new axes, and provided we replace (x, y, z) by (x', y', z'). In the transformations of pre-relativistic physics the time was naïvely assumed to be an invariant. Unlike the co-ordinates (x, y, z) it suffered no change in the passage from one system of co-ordinates to another fixed relatively to it. The reason why the expression **(34·5)** represents the X component of the force per unit volume in any system of reference is of course that both are the X components of vectors and are therefore transformed or changed according to the same rule on passing from one system of co-ordinates to another.

The statement made above, namely, that in pre-relativistic physics the equations describing physical phenomena are equations connecting vector and tensor quantities and that they have consequently the same form in all systems of reference which are fixed relatively to one another in Euclidean space is in reality a statement of a very restricted form of the **principle of relativity.** What is new in the theory of relativity is not so much the introduction of this principle as its recognition and extension.

Now that we have met with strong grounds for the view that we can render a still better and more coherent account of physical phenomena by building up a more comprehensive 4-dimensional continuum, in which space and time are woven into a single whole, we are almost forced to imitate as closely as we can, with the new continuum, the old-fashioned procedure with the old Euclidean one. That is to say, we are now forced to regard the equations describing physical phenomena as vector and tensor equations in the new continuum and we must lay down that they shall have the same form in all G co-ordinate systems which are ' fixed ' [1] relatively to one another in Galilean space-time. This means that **the equations describing physical phenomena are to have the same form in all E reference systems which are moving relatively to one another with constant velocities.** This is the **principle of relativity** as enunciated by Einstein in 1905. We shall, however, employ this term for the principle as extended still further by Einstein, at a later date, to apply to all reference systems and shall use the term **special principle of relativity** (also adopted by

[1] The term ' fixed ' is used here for want of a better one. When used earlier with reference to co-ordinate systems ' fixed ' relatively to one another in Euclidean space its meaning was clear enough. The relationship between such co-ordinate systems has a simple mathematical description: namely, the equations of transformation. The term ' fixed ' as now used has the same meaning for Galilean co-ordinate systems. Cf. **34·1** where the meanings of ' straight line ', ' parallel ', ' rotation ', etc., in a Galilean continuum are explained.

Einstein) for the somewhat restricted form of it which we are now studying.

We may say that the special principle of relativity affirms that the equations describing physical phenomena, whether they are referred to one set of rectangular E co-ordinates or to any other moving relatively to it with a constant velocity of translation, are identical in their form. The term ' relativity ' as originally used had special reference to this fact. It is worthy of notice that the special principle of relativity applies completely to the elementary equations of Newtonian dynamics in its old-fashioned form, i.e. without giving up the older view of a Euclidean space and a separate invariant time. Consider, for example, the equations representing the force exerted on a particle and carry out the transformation from a co-ordinate system (X, Y, Z) to another (X', Y', Z') using the old-fashioned equations of transformation **(34·25).** The mass, m, of the particle and also the time are invariants,[1] so that md^2x/dt^2 is in fact identical with md^2x'/dt'^2. The association of the special principle of relativity with a 4-dimensional continuum of space and time is necessary to enable us to extend it to the electromagnetic field equations and the propagation of light, which would be impossible if we were to retain the older view of the independence of space and time. It is true that Einstein did not, in his first paper, employ the notion of a 4-dimensional continuum, which appears to have occurred first to Minkowski, but he used the negative result of Michelson's experiment as an axiom and this, combined with his special principle of relativity, amounted to the same thing.

§ 34·6. The Aether

Einstein, as indicated above, was not led to his special theory of relativity by the line of thought set out in this introduction. He was impressed by the fact that Maxwellian electrodynamics, as originally interpreted and applied, leads to asymmetries which do not show themselves in any observations, and he pointed out in illustration that, in the case of the relative motion of a conductor and a magnet, the observed effects depend *only on the relative motion*, and that it does not matter in any way whether the conductor is ' at rest ' and the magnet ' in motion ' or whether the conductor is ' in motion ' and the magnet ' at rest '. According to the old interpretation the magnetic lines of induction are in motion through the aether in the former case, giving rise to an electric field and a consequent induced current; while in

[1] In Newtonian dynamics.

the latter case the lines are at rest in the aether; there can be, it would seem, no electric field produced; but nevertheless an induced current is observed. Thus, while the two cases are observationally identical, the old theory, or at any rate the old interpretation of Maxwell's theory, distinguishes one from the other. Reflections such as these, combined with the problem raised by Michelson's negative result **(§ 33·4)**—a result difficult or impossible to reconcile with the view that light is propagated through an aether—led Einstein to ignore the aether altogether and to adopt as a fundamental principle that relative motions alone determine the nature of the phenomena observed.

The abolition of the aether renders the view of the nature of electromagnetic waves, that we have expressly or tacitly held up to this point, untenable. The transport of energy which is associated with them—through free space as well as through material media—requires that we must either (*a*) assume a medium (aether) in which the energy is seated, or (*b*) adopt an entirely new interpretation of the significance of Maxwellian electromagnetic theory and of electromagnetic waves. The latter alternative is forced on us not only by the theory of relativity, but by the observed facts of photoelectricity and of the emission of light (spectra) and their theoretical explanation by Planck's quantum theory, more especially by its modern developments, among which may be mentioned here the **wave mechanics** of Louis de Broglie and Erwin Schroedinger. Roughly speaking, we have to revert to Newton's views about the nature of light and regard it as made of ' rays ', as Newton termed them : something resembling small particles or corpuscles—they are called **photons**—with each of which is associated some kind of *phenomène periodique* (Newton's ' fits '). The mathematical expression of electromagnetic theory, including electromagnetic waves, remains intact, with a new significance the description of which we shall leave till a later chapter.

CHAPTER II

THE SPECIAL THEORY OF RELATIVITY

§ 34·7. Vector and Tensor Analysis in the Galilean Continuum

IN this and the following section we shall continue to use orthogonal (rectangular) co-ordinate systems, i.e. co-ordinate systems in which the square of the interval ds has the form

$$ds^2 = dx^2 + dy^2 + dz^2 + dw^2,$$

or

$$ds^2 = dx^2 + dy^2 + dz^2 - c^2\,dt^2.$$

It is, of course, open to us to use, for example, polar co-ordinates for spacial determinations, in which case the square of the interval ds will be expressed by

$$ds^2 = dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2 - c^2\,dt^2,$$

and we shall enlarge the scope of tensor analysis in **§ 36·2** so that it will be entirely independent of the choice we make of co-ordinate systems. Meanwhile, however, we shall confine our studies to a somewhat narrower field.

The typical vector in the Galilean continuum, just as in the Euclidean one (cf. **§§ 2·2** and **34·1**), is the **displacement,**

$$(dx,\ dy,\ dz,\ dw),$$

the components of which are referred to axes X, Y, Z, W. It is easy to see that the components

$$(dx',\ dy',\ dz',\ dw')$$

of this vector, referred to the axes X', Y', Z', W', are related to those referred to the former system by equations of the form:

$$\left.\begin{aligned} dx' &= \alpha_{11}\,dx + \alpha_{12}\,dy + \alpha_{13}\,dz + \alpha_{14}\,dw,\\ dy' &= \alpha_{21}\,dx + \alpha_{22}\,dy + \alpha_{23}\,dz + \alpha_{24}\,dw,\\ dz' &= \alpha_{31}\,dx + \alpha_{32}\,dy + \alpha_{33}\,dz + \alpha_{34}\,dw,\\ dw' &= \alpha_{41}\,dx + \alpha_{42}\,dy + \alpha_{43}\,dz + \alpha_{44}\,dw, \end{aligned}\right\} \quad . \qquad (34\cdot7)$$

where the coefficients, α, are constants. This in fact is just as in Euclidean space—the only difference is the extension to four

dimensions. The Lorentz transformation, **(34·21)** and **(34·211)**, is the particular case where

$$\alpha_{11} = \cos\phi,\ \alpha_{14} = -\sin\phi,\ \alpha_{22} = 1$$
$$\alpha_{33} = 1,\ \alpha_{41} = \sin\phi,\ \alpha_{44} = \cos\phi,$$

and the remaining α's vanish.

The invariance of the expression **(34·1)** under the transformation **(34·7)** leads necessarily to relationships of which the following are typical examples:

$$\alpha_{11}^2 + \alpha_{21}^2 + \alpha_{31}^2 + \alpha_{41}^2 = 1,$$
$$\alpha_{11}\alpha_{12} + \alpha_{21}\alpha_{22} + \alpha_{31}\alpha_{32} + \alpha_{41}\alpha_{42} = 0,$$

(cf. **§ 2·2**). These relationships may be more compactly expressed as follows:

$$\sum_s \alpha_{rs}\alpha_{ps} = \sum_s \alpha_{sr}\alpha_{sp} = \begin{cases} 1, & r = p, \\ 0, & r \neq p, \end{cases}$$

which we shall often abbreviate by the use of the summation convention, that is to say we shall drop the symbol $\sum_s$, and leave the repetition of the s to perform its function, thus:

$$\alpha_{rs}\alpha_{ps} = \alpha_{sr}\alpha_{sp} = \begin{cases} 1, & r = p, \\ 0, & r \neq p. \end{cases} \quad . \quad . \quad . \quad \textbf{(34·71)}$$

To these relationships may be added another, which in fact emerges from them, namely

$$\begin{vmatrix} \alpha_{11}, & \alpha_{12}, & \alpha_{13}, & \alpha_{14} \\ \alpha_{21}, & \alpha_{22}, & \alpha_{23}, & \alpha_{24} \\ \alpha_{31}, & \alpha_{32}, & \alpha_{33}, & \alpha_{34} \\ \alpha_{41}, & \alpha_{42}, & \alpha_{43}, & \alpha_{44} \end{vmatrix} = 1 \quad . \quad . \quad . \quad \textbf{(34·72)}$$

It can be established by applying the rule for multiplying determinants to the product

$$\begin{vmatrix} \alpha_{11}, & \alpha_{12}, & \alpha_{13}, & \alpha_{14} \\ \alpha_{21}, & \alpha_{22}, & \alpha_{23}, & \alpha_{24} \\ \alpha_{31}, & \alpha_{32}, & \alpha_{33}, & \alpha_{34} \\ \alpha_{41}, & \alpha_{42}, & \alpha_{43}, & \alpha_{44} \end{vmatrix} \times \begin{vmatrix} \alpha_{11}, & \alpha_{21}, & \alpha_{31}, & \alpha_{41} \\ \alpha_{12}, & \alpha_{22}, & \alpha_{32}, & \alpha_{42} \\ \alpha_{13}, & \alpha_{23}, & \alpha_{33}, & \alpha_{43} \\ \alpha_{14}, & \alpha_{24}, & \alpha_{34}, & \alpha_{44} \end{vmatrix}$$

and making use of **(34·71)**. This product, which is obviously equal to the square of the determinant **(34·72)**, turns out to be $+1$. This makes the determinant itself equal to $+1$ or -1, and it is easy to demonstrate that of the two signs it is the positive one which must be adopted.

Any set of four quantities

$$A_x,\ A_y,\ A_z,\ A_w,$$

which obey the law of transformation **(34·7)**, i.e. the law:

$$\begin{aligned} A_x' &= \alpha_{11}A_x + \alpha_{12}A_y + \alpha_{13}A_z + \alpha_{14}A_w, \\ A_y' &= \alpha_{21}A_x + \alpha_{22}A_y + \alpha_{23}A_z + \alpha_{24}A_w, \\ A_z' &= \alpha_{31}A_x + \alpha_{32}A_y + \alpha_{33}A_z + \alpha_{34}A_w, \\ A_w' &= \alpha_{41}A_x + \alpha_{42}A_y + \alpha_{43}A_z + \alpha_{44}A_w, \end{aligned}$$

is defined to be a **vector** or **tensor of rank one.** Any single quantity which remains invariant under the transformation from X, Y, Z, W, to X', Y', Z', W', will be termed a **scalar** quantity or **tensor of rank zero.**

A **tensor of rank two** is a set of 16 quantities (its components) which we may represent by

$$A_{xx},\ A_{xy},\ A_{xz},\ A_{xw},$$
$$A_{yx},\ A_{yy},\ \text{etc.},$$

each of which is subject to the same law of transformation as the corresponding product of the components of two vectors ; that is to say the formula for calculating the component A_{xy}' (for example) in the X', Y', Z', W' system of co-ordinates from the components A_{xx}, A_{xy}, etc., in the X, Y, Z, W system is identical with that for calculating the product $R_x'S_y'$ from the products R_xS_x, R_xS_y, etc., **R** and **S** being any two vectors.[1]

If we distinguish components by numerical subscripts, instead of the letters x, y, z, w, and make use of the summation convention, we may give the foregoing definitions and the corresponding transformation formulae with more brevity and precision. For a vector, **A,** for example, we have the four equations of transformation:

$$A_\mu' = \alpha_{\mu m}A_m. \quad . \quad . \quad . \quad . \quad \textbf{(34·73)}$$

For a tensor of rank two we have

$$A_{\mu\nu}' = \alpha_{\mu m}\alpha_{\nu n}A_{mn} \quad . \quad . \quad . \quad . \quad \textbf{(34·74)}$$

and so on.

It will be seen that a **tensor of zero rank (scalar)** has 1 component, that is to say, 4^0 components ; a **tensor of rank one (vector)** has 4^1 components ; a **tensor of rank two** has 4^2 ; and so on further.

The scalar product,

$$A_mB_m,$$

(the summation convention is now being used) is an invariant, since

$$A_\mu'B_\nu' = \alpha_{\mu m}\alpha_{\nu n}A_mB_n,$$

and consequently

$$A_\mu'B_\mu' = \alpha_{\mu m}\alpha_{\mu n}A_mB_n.$$

[1] Cf. §§ **2·2, 2·3** and **19·2.**

When the summation with respect to μ is carried out in $\alpha_{\mu m}\alpha_{\mu n}$ on the right, we get zero, except when $n = m$ **(34·71).** Therefore

$$A_\mu' B_\mu' = A_m B_m.$$

The product $A_{mn}B_m$ is a tensor of rank one (vector), since

$$A_{\mu\nu}' B_\tau' = \alpha_{\mu m}\alpha_{\nu n}\alpha_{\tau t} A_{mn} B_t,$$

and therefore

$$A_{\mu\nu}' B_\mu' = \alpha_{\mu m}\alpha_{\nu n}\alpha_{\mu t} A_{mn} B_t.$$

But the sum $\alpha_{\mu m}\alpha_{\mu t}$ is zero except when $t = m$, in which case it is equal to unity. Consequently

$$A_{\mu\nu}' B_\mu' = \alpha_{\nu n} A_{mn} B_m,$$

which might be written

$$C_\nu' = \alpha_{\nu n} C_n,$$

so that C_ν or C_ν', that is to say $A_{mn}B_m$ or $A_{\mu\nu}'B_\mu'$, is a vector.

§ 34·8. Covariant and Contravariant Tensors

We shall here make a distinction which will be of great importance when we pass beyond the special theory of relativity, but which is a purely formal one in a Euclidean or Galilean continuum when we employ orthogonal co-ordinates. It is easy to show that the gradient of a scalar quantity, ϕ, is a vector. For example

$$\frac{\partial\phi}{\partial x'} = \frac{\partial\phi}{\partial x}\frac{\partial x}{\partial x'} + \frac{\partial\phi}{\partial y}\frac{\partial y}{\partial x'} + \frac{\partial\phi}{\partial z}\frac{\partial z}{\partial x'} + \frac{\partial\phi}{\partial w}\frac{\partial w}{\partial x'}, \quad . \qquad \textbf{(34·8)}$$

and since

$$x = \alpha_{11}x' + \alpha_{21}y' + \alpha_{31}z' + \alpha_{41}w',$$
$$y = \alpha_{12}x' + \alpha_{22}y' + \alpha_{32}z' + \alpha_{42}w',$$

and so on, we must have

$$\left.\begin{aligned} \frac{\partial x}{\partial x'} &= \alpha_{11}, \\ \frac{dy}{\partial x'} &= \alpha_{12}, \\ \frac{\partial z}{\partial x'} &= \alpha_{13}, \\ \frac{\partial w}{\partial x'} &= \alpha_{14}. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \qquad \textbf{(34·801)}$$

Consequently

$$\frac{\partial\phi}{\partial x'} = \alpha_{11}\frac{\partial\phi}{\partial x} + \alpha_{12}\frac{\partial\phi}{\partial y} + \alpha_{13}\frac{\partial\phi}{\partial z} + \alpha_{14}\frac{\partial\phi}{\partial w},$$

which is the equation of transformation of a vector.

For the typical vector, on the other hand, we have

$$dx' = \frac{\partial x'}{\partial x}dx + \frac{\partial x'}{\partial y}dy + \frac{\partial x'}{\partial z}dz + \frac{\partial x'}{\partial w}dw, \quad . \quad \textbf{(34·81)}$$

and the coefficients α were so defined that

$$\left.\begin{aligned} \frac{\partial x'}{\partial x} &= \alpha_{11}, \\ \frac{\partial x'}{dy} &= \alpha_{12}, \\ \frac{\partial x'}{\partial z} &= \alpha_{13}, \\ \frac{\partial x'}{\partial w} &= \alpha_{14}. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad \textbf{(34·811)}$$

The comparison of **(34·8)** and **(34·81)** reveals an obvious formal difference. In the former case we find the coefficients:

$$\frac{\partial x}{\partial x'}, \frac{dy}{\partial x'}, \frac{\partial z}{\partial x'}, \frac{\partial w}{\partial x'}, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(34·82)}$$

in the latter:

$$\frac{\partial x'}{\partial x}, \frac{\partial x'}{dy}, \frac{\partial x'}{\partial z}, \frac{\partial x'}{\partial w}. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(34·821)}$$

And so we have two groups of vectors: one group conforming to the same law of transformation as **(34·8)** and the other to the same law as **(34·81)**. The distinction is purely formal in the Galilean continuum, when we use orthogonal co-ordinates, because in it the coefficients **(34·821)** are the constants α_{11}, α_{12}, α_{13}, α_{14} and are necessarily identical with the coefficients **(34·82)**. Vectors of the former kind are called **covariant** to distinguish them from vectors of the latter kind which are called **contravariant.** We have of course covariant and contravariant tensors of all ranks, since the definitions of tensors of higher ranks are based on those of vectors. It is usual to distinguish covariance and contravariance by the positions of the indices which mark the components. For example, the components of a covariant vector are usually written:

$$A_1, A_2, A_3, A_4,$$

while those of a contravariant one are written:

$$A^1, A^2, A^3, A^4.$$

A covariant vector obeys the law of transformation **(34·8)**, namely

$$A_\mu' = \frac{\partial x^1}{\partial x'^\mu}A_1 + \frac{\partial x^2}{\partial x'^\mu}A_2 + \frac{\partial x^3}{\partial x'^\mu}A_3 + \frac{\partial x^4}{\partial x'^\mu}A_4,$$

or, if we employ the summation convention,

$$A_\mu{}' = \frac{\partial x^m}{\partial x'^\mu} A_m. \quad . \quad . \quad . \quad . \quad . \quad (34\cdot83)$$

The law of transformation for a contravariant tensor is (34·81)

$$A'^\mu = \frac{\partial x'^\mu}{\partial x^m} A^m. \quad . \quad . \quad . \quad . \quad . \quad (34\cdot84)$$

The equations of transformation for a covariant tensor A_{mn} of the second rank are

$$A_{\mu\nu}{}' = \frac{\partial x^m}{\partial x'^\mu} \frac{\partial x^n}{\partial x'^\nu} A_{mn}, \quad . \quad . \quad . \quad (34\cdot841)$$

and those for a contravariant tensor A^{mn} are

$$A'^{\mu\nu} = \frac{\partial x'^\mu}{\partial x^m} \frac{\partial x'^\nu}{\partial x^n} A^{mn}. \quad . \quad . \quad . \quad (34\cdot842)$$

Lastly, we may give the following illustration of a **mixed tensor :**

$$A'^{\mu\nu}_{\omega} = \frac{\partial x'^\mu}{\partial x^m} \frac{\partial x'^\nu}{\partial x^n} \frac{\partial x^o}{\partial x'^\omega} A^{mn}_{o}. \quad . \quad . \quad . \quad (34\cdot85)$$

Suppose that, in the last equation, we replace the covariant suffix, ω, by ν and carry out the summation which the notation will now imply,

$$A'^{\mu\nu}_{\nu} = \frac{\partial x'^\mu}{\partial x^m} \frac{\partial x'^\nu}{\partial x^n} \frac{\partial x^o}{\partial x'^\nu} A^{mn}_{o}.$$

Now

$$\frac{\partial x'^\nu}{\partial x^n} \frac{\partial x^o}{\partial x'^\nu} = \frac{\partial x^o}{\partial x^n}$$

which is equal to unity when $0 = n$ and zero when $0 \neq n$. Hence

$$A'^{\mu\nu}_{\nu} = \frac{\partial x'^\mu}{\partial x^m} A^{mn}_{n}.$$

Thus

$$A'^{\mu\nu}_{\nu} \text{ or } A^{mn}_{n}$$

represents a vector, in this illustration a contravariant one, and we might represent it by

$$A'^\mu \text{ or } A^m.$$

This process, whereby we obtain one vector (or tensor) from another, is called **contraction.**

In equations **(34·01)** and **(34·02)** the notation has already anticipated the contravariant and covariant character which further investigation will lead us to assign to the respective field tensors involved.

Suppose we are given that a set of four quantities $A(m)$, $m = 1, 2, 3, 4$—the index, m, has been given a non-committal

position to avoid prejudging the question as to whether they constitute a vector of either kind—and suppose we are given, further, that

$$A_{(m)}B^m$$

is invariant, B^m being *any* contravariant vector. We inquire about $A_{(m)}$. We are given then that

$$A'_{(\mu)}B'_\mu = A_{(m)}B^m,$$

and since B^m is a contravariant vector

$$B^m = \frac{\partial x^m}{\partial x'^\mu}B'^\mu,$$

and therefore

$$A'_{(\mu)}B'^\mu = \frac{\partial x^m}{\partial x'^\mu}A_{(m)}B'^\mu.$$

But the components B'^μ are quite arbitrary and hence

$$A'_{(\mu)} = \frac{\partial x^m}{\partial x'^\mu}A_{(m)};$$

so that $A_{(m)}$ is a covariant vector A_m.

As another example, let us take the set of 16 quantities $A_{(m,\,n)}$, $m = 1, 2, 3, 4$ and $n = 1, 2, 3, 4$. If we are given, for example, that

$$A_{(m,\,n)}B_n$$

is a contravariant vector, say C^m, when B_n is *any* covariant vector, then since

$$C^m = \frac{\partial x^m}{\partial x'^\mu}C'^\mu,$$

we must have

$$A_{(m,\,n)}B_n = \frac{\partial x^m}{\partial x'^\mu}A'_{(\mu\nu)}B_\nu';$$

and we are given that

$$B_\nu' = \frac{\partial x^n}{\partial x'^\nu}B_n,$$

consequently

$$A_{(m,\,n)}B_n = \frac{\partial x^m}{dx'^\mu}\cdot\frac{\partial x^n}{\partial x'^\nu}A'_{(\mu,\,\nu)}B_n.$$

The components B_n are arbitrary and therefore

$$A_{(m,\,n)} = \frac{\partial x^m}{\partial x'^\mu}\frac{\partial x^n}{\partial x'^\nu}A'_{(\mu,\,\nu)}.$$

Thus $A_{(m,\,n)}$ is a contravariant tensor $A^{m,\,n}$.

All propositions of this kind are included in the following general statement, the correctness of which can easily be demonstrated:

We are given a set of quantities $A(a, b, \ldots e, f, \ldots i, k, \ldots m, n)$ with the property that

$$A(a, b \ldots e, f \ldots i, k \ldots m, n)\, B^{a, b \ldots r, s}_{e, f \ldots t, x}$$

is a tensor, namely,

$$C^{r, s \ldots i, k}_{t, x \ldots m, n},$$

$B^{a, b \ldots r, s}_{e, f \ldots t, x}$ being an *arbitrary* tensor of the kind which the notation indicates. The set of quantities **A** constitutes a tensor

$$A^{e, f \ldots i, k}_{a, b \ldots m, n}.$$

When the components of a vector (or tensor) are all multiplied by the same invariant, ϕ, it remains a vector (or tensor) and its character (covariance or contravariance) is unmodified. The typical covariant vector,

$$\left(\frac{\partial\phi}{\partial x^1}, \frac{\partial\phi}{\partial x^2}, \frac{\partial\phi}{\partial x^3}, \frac{\partial\phi}{\partial x^4}\right),$$

has the *form* of such a product and consequently (since ϕ is invariant) the law of transformation of the operations

$$\left(\frac{\partial}{\partial x^1}, \frac{\partial}{\partial x^2}, \frac{\partial}{\partial x^3}, \frac{\partial}{\partial x^4}\right) \quad . \quad . \quad . \quad . \quad \textbf{(34·86)}$$

is that obeyed by the components of a covariant vector and we may appropriately call it a **symbolic vector.**[1] It is of course necessarily covariant.

The reader will have no difficulty in demonstrating, for example, that the set of quantities

$$\frac{\partial A^{mn}}{\partial x^n} = C^m$$

constitutes a contravariant vector,[2] or that

$$\frac{\partial A^m}{\partial x^m}$$

is an invariant.

The general statement including all the propositions of this type is:

$$\frac{\partial A^{a, b \ldots e, f \ldots}_{c, d \ldots}}{\partial x^a\, \partial x^b \ldots \partial x^i\, \partial x^k \ldots} \text{ is a tensor}$$

which may be written $C^{e, f \ldots}_{c, d \ldots i, k \ldots}$

It is important to bear in mind that this and similar statements cannot be true when $\partial x'^1/\partial x^1$, $\partial x'^1/\partial x^2$, etc., **are not constants ;**

[1] Cf. **§ 2·4.**

[2] It should be remembered that we have orthogonal co-ordinates in mind.

as for example when we use in Euclidean space or in Galilean space-time other types of co-ordinate systems, e.g. polar co-ordinates, r, θ, ϕ, and define the typical contravariant vector as

$$(dr,\ d\theta,\ d\phi),$$

or

$$(dr,\ d\theta,\ d\phi,\ dw).$$

A simple example will make this clear. Consider the set of quantities, $\partial A^m/\partial x^n$. The equation of transformation for any one of these may be obtained from

$$\frac{\partial A^m}{\partial x^n} = \frac{\partial}{\partial x^n}\left\{\frac{\partial x^m}{\partial x'^\mu}A'^\mu\right\}.$$

Now **when the coefficients** $\partial x^m/\partial x'^\mu$ **are constants, and only then,** we may write this in the form :

$$\frac{\partial A^m}{\partial x^n} = \frac{\partial x^m}{\partial x'^\mu}\frac{\partial A'^\mu}{\partial x^n},$$

or

$$\frac{\partial A^m}{\partial x^n} = \frac{\partial x^m}{\partial x'^\mu}\frac{\partial x'^\nu}{\partial x^n}\frac{\partial A'^\mu}{\partial x'^\nu}$$

and we see that we have a tensor (in this illustration a mixed one) of rank two.

§ 34·9. The Velocity Vector in Galilean Space-Time

The components

$$dx/dt,\ dy/dt,\ dz/dt,$$

of the velocity (of a particle, for example) in Euclidean space may not be identified with the X, Y, and Z components respectively of a vector in Galilean space-time; since they do not obey the appropriate laws of transformation, namely the laws obeyed by dx, dy, and dz, the corresponding components of the displacement of a particle, which together with dw ($= ic\,dt$ where dt is the time for the displacement to occur), constitute the typical vector in Galilean space-time. This is due to the fact that dt is not an invariant in the Galilean continuum; but changes as we pass from one system of co-ordinates to another in a way which has already been described. If we wish to find a vector which will correspond as closely as possible to a velocity in Euclidean space, we must search for a differential—let us call it $d\tau$—which, while being an invariant, will in general differ very little from the corresponding dt. The expression

$$ds^2 = dx^2 + dy^2 + dz^2 - c^2\,dt^2$$

is an important invariant. If we write it in the form

$$dx^2 + dy^2 + dz^2 - c^2\,dt^2 = -\,c^2\,d\tau^2, \quad . \quad . \quad \textbf{(34·9)}$$

we see that $d\tau^2$, and consequently $d\tau$, is an invariant, and since

$$d\tau^2/dt^2 = 1 - (dx^2 + dy^2 + dz^2)/(c^2\,dt^2),$$

or

$$d\tau^2/dt^2 = 1 - v^2/c^2,$$

where v is the velocity of the particle as ordinarily understood, we see that the ratio $d\tau/dt$ differs from unity by a second order small quantity when v/c is small. So that, when this condition (i.e. $v/c \ll 1$) holds, $d\tau$ and dt are to all intents and purposes identical. Therefore

$$dx/d\tau,\ dy/d\tau,\ dz/d\tau,\ dw/d\tau,\ \ldots \quad \textbf{(34·91)}$$

is such a vector as we require. We may note that

$$dt = (1 - v^2/c^2)^{-1/2}\,d\tau,$$

or

$$dt = \gamma\,d\tau, \quad \ldots \quad \textbf{(34·92)}$$

where γ is the familiar function of the velocity, v, we have met with in the Lorentz transformation and elsewhere.

It should be observed that as v approaches c as a limit the components of the Galilean velocity **(34·91)** approach the limit ∞, since $d\tau/dt$ approaches the limit 0 in this case. The invariant, τ, is called the **proper time** (German: *Eigenzeit*). It is in fact the time referred to a co-ordinate system in which the particle whose velocity we are discussing is at rest. It is important that this should be kept in mind. It means that **at a given instant the proper time has generally different values for different particles.**

§ 35. Mass, Momentum and Energy

We shall proceed further by endeavouring to extend the old-fashioned Newtonian particle dynamics just as we have extended our geometry from the old Euclidean 3-dimensional type to the 4-dimensional Galilean geometry. Let us therefore write for the equations of motion of a particle

$$\left.\begin{aligned} m_0\frac{d^2x}{d\tau^2} &= \mathscr{F}_x,\\ m_0\frac{d^2y}{d\tau^2} &= \mathscr{F}_y,\\ m_0\frac{d^2z}{d\tau^2} &= \mathscr{F}_z,\\ m_0\frac{d^2w}{d\tau^2} &= \mathscr{F}_w, \end{aligned}\right\} \quad \ldots \quad \textbf{(35)}$$

where m_0 is a constant characteristic of the particle and invariant under co-ordinate transformations and $(\mathscr{F}_x, \mathscr{F}_y, \mathscr{F}_z, \mathscr{F}_w)$

is a 4-dimensional vector. When the velocity of the particle is small (compared with c) the new dynamics we are developing must coincide with Newtonian dynamics, and we have already seen **(§ 34·9)** that, when $v/c \ll 1$, τ becomes practically identical with t. In this case we may therefore identify $\mathscr{F}_x$, $\mathscr{F}_y$, and $\mathscr{F}_z$ with the components of the Newtonian force acting on the particle and the constant, m_0, with its mass. When the velocity is not small, so that $d\tau$ is appreciably different from dt, we shall regard **(35)** as the equations defining the vector $(\mathscr{F}_x, \mathscr{F}_y, \mathscr{F}_z, \mathscr{F}_w)$. It is convenient to have a name for it and we shall call it the **Minkowskian force** after H. Minkowski who was the first to recognize that space and time constitute a single continuum.

In the next place we shall adopt the **principle of conservation of momentum** and by the **momentum** of a particle we shall understand the vector

$$\left(m_0\frac{dx}{d\tau},\ m_0\frac{dy}{d\tau},\ m_0\frac{dz}{d\tau},\ m_0\frac{dw}{d\tau}\right) \quad . \quad . \quad . \quad \textbf{(35·01)}$$

Thus we shall lay down for a number of particles, free from all interference except that due to collisions with one another, that

$$\left.\begin{aligned} \Sigma m_0\frac{dx}{d\tau} &= \alpha, \\ \Sigma m_0\frac{dy}{d\tau} &= \beta, \\ \Sigma m_0\frac{dz}{d\tau} &= \gamma, \\ \Sigma m_0\frac{dw}{d\tau} &= \delta, \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad \textbf{(35·02)}$$

α, β, γ, and δ being constants. It is easy to see that if these equations are true in one system of co-ordinates they must be true in any system of (rectangular) co-ordinates. For example

$$\begin{aligned} \Sigma m_0\frac{dx'}{d\tau} &= \alpha_{11}\Sigma m_0\frac{dx}{d\tau} + \alpha_{12}\Sigma m_0\frac{dy}{d\tau} + \alpha_{13}\Sigma m_0\frac{dz}{d\tau} + \alpha_{14}\Sigma m_0\frac{dw}{d\tau} \\ &= \alpha_{11}\alpha + \alpha_{12}\beta + \alpha_{13}\gamma + \alpha_{14}\delta \\ &= \alpha', \text{ a constant.} \end{aligned}$$

Similarly

$$\Sigma m_0\frac{dy'}{d\tau} = \beta',$$

$$\Sigma m_0\frac{dz'}{d\tau} = \gamma',$$

$$\Sigma m_0\frac{dw'}{d\tau} = \delta'.$$

Let us now multiply and divide each term in the summations, Σ, of **(35·02)** by the appropriate factor γ. The constant, α, for example, is the sum of a number of products $m_0\,dx/d\tau$, each one belonging to a separate particle. For each particle there is a definite value of v^2, namely $(dx^2 + dy^2 + dz^2)/dt^2$, and therefore a definite value of $(1 - v^2/c^2)^{-1/2}$ which we represent by the letter γ. So we give each term the form

$$\gamma m_0 \frac{dx}{\gamma d\tau} = \gamma m_0 \frac{dx}{dt}$$

since

$$\gamma d\tau = dt.$$

The equations **(35·02)** thus become, if we represent γm_0 by m

$$\left.\begin{aligned} \Sigma m\frac{dx}{dt} &= \alpha, \\ \Sigma m\frac{dy}{dt} &= \beta, \\ \Sigma m\frac{dz}{dt} &= \gamma, \\ \Sigma m\frac{dw}{dt} &= \delta. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad \textbf{(35·021)}$$

The velocity components dx/dt, dy/dt, and dz/dt are velocity components in the ordinary and familiar sense of the term; but they are not the corresponding components of a vector in Galilean space-time; but $m\,dx/dt$ is of course the X component of such a vector since it is equal to $m_0\,dx/d\tau$ and thus equal to dx multiplied by an invariant. The factor m is defined to be the **mass** of the particle. It varies with the velocity of the particle according to the law

$$m = \gamma m_0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(35·03)}$$

or

$$m = m_0(1 - v^2/c^2)^{-1/2}$$

and becomes identical with m_0 when the velocity v becomes zero.

It is instructive to examine the last of the equations **(35·021)**. Since $dw = ic\,dt$ it is equivalent to

$$\Sigma m\,ic = \text{a constant}$$

or

$$\Sigma m = \text{a constant},$$

and we see that mass, as we have defined it, is conserved. The **law of conservation of momentum** in Galilean space-time includes therefore not only the law of conservation of momentum in its old-fashioned form, but the **law of conservation of mass** as well.

In the next place it may be noted that the sum of the squares of the four momentum components of a particle, namely

$$m_0{}^2\frac{dx^2}{d\tau^2} + m_0{}^2\frac{dy^2}{d\tau^2} + m_0{}^2\frac{dz^2}{d\tau^2} + m_0{}^2\frac{dw^2}{d\tau^2},$$

is a constant. It is equal to $m_0{}^2\, ds^2/d\tau^2$ which, by the definition of $d\tau^2$, is equal to $-\, m_0{}^2c^2$. The absolute value of the momentum of a particle, in the widened Galilean sense, is $m_0 ic$ and constant. We may therefore say that the Minkowskian force acting on a particle and its velocity, $(dx/d\tau, dy/d\tau, dz/d\tau, dw/d\tau)$ are orthogonal. This is otherwise evident when we multiply **(35)** by $dx/d\tau$, $dy/d\tau$, $dz/d\tau$, and $dw/d\tau$ respectively, and add. We thus get

$$\begin{aligned}\mathscr{F}_x\frac{dx}{d\tau} + \mathscr{F}_y\frac{dy}{d\tau} + \mathscr{F}_z\frac{dz}{d\tau} + \mathscr{F}_w\frac{dw}{d\tau}&\\ = \frac{m_0}{2}\,\frac{d}{d\tau}\left\{\left(\frac{dx}{d\tau}\right)^2 + \left(\frac{dy}{d\tau}\right)^2 + \left(\frac{dz}{d\tau}\right)^2 + \left(\frac{dw}{d\tau}\right)^2\right\}&\\ = \frac{m_0}{2}\,\frac{d}{d\tau}\{-c^2\} \quad . \quad . \quad . \quad . \quad . \quad \textbf{(35·04)}&\\ = 0&\end{aligned}$$

Equation **(35·04)** therefore affirms that the 4-dimensional scalar product **((Minkowskian force, velocity))** is always zero.

Turning back to any one of the four equations which express the principle of conservation of momentum, for example to

$$\Sigma m_0\frac{dx}{d\tau} = \text{constant},$$

we infer that

$$\sum\frac{d}{dt}\left\{m_0\frac{dx}{d\tau}\right\} = 0,$$

or

$$\sum\frac{d}{dt}\left\{m\frac{dx}{dt}\right\} = 0.$$

We may write this in the form

$$\Sigma F_x = 0,$$

and regard F_x as the X component of a somewhat differently defined force, namely

$$\left.\begin{aligned}\frac{d}{dt}\left(m\frac{dx}{dt}\right) &= F_x,\\ \frac{d}{dt}\left(m\frac{dy}{dt}\right) &= F_y,\\ \frac{d}{dt}\left(m\frac{dz}{dt}\right) &= F_z,\\ \frac{d}{dt}\left(m\frac{dw}{dt}\right) &= F_w.\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad \textbf{(35·05)}$$

The X, Y, and Z components of this force resemble very closely those of the force as it appears in Newtonian and Hamiltonian dynamics. In one respect, however, we have passed out beyond the scope of Newtonian dynamics. The mass of a particle is now a function of its velocity—the same function in fact that we were led to by the considerations of **§§ 27·1** and **27·2.** It is convenient to call the set of four quantities (F_x, F_y, F_z, F_w) the **Newtonian force ; but it should be noted that it is not a vector in our 4-dimensional continuum.**

We may note that the law of action and reaction (Newton's third law) is obeyed by the Newtonian force as it has been defined, but not for the Minkowskian force.

We observe next that

$$\mathscr{F}_x/F_x = \mathscr{F}_y/F_y = \mathscr{F}_z/F_z = \mathscr{F}_w/F_w = \gamma. \quad . \quad \textbf{(35·06)}$$

Remembering this and also that

$$v_x/(dx/d\tau) = v_y/(dy/d\tau) = v_z/(dz/d\tau) = v_w/(dw/d\tau) = 1/\gamma, \quad . \quad . \quad . \quad \textbf{(35·07)}$$

we get from **(35·4)**

$$F_x v_x + F_y v_y + F_z v_z + F_w v_w = 0. \quad . \quad . \quad \textbf{(35·08)}$$

The rate at which work is done on the particle we shall naturally identify with the (E) scalar product of the Newtonian force and velocity, namely

$$F_x v_x + F_y v_y + F_z v_z ;$$

so that the rate at which the energy of the particle increases is

$$-F_w v_w \text{ or } -F_w dw/dt \text{ or } -F_w ic. \quad . \quad \textbf{(35·085)}$$

In the case of a system of particles subject only to their mutual interactions

$$\Sigma F_w = 0,$$

as we have seen ; hence

$$\Sigma - F_w ic = 0.$$

Otherwise expressed, the rate of increase of the energy of the system is zero, and the **principle of conservation of energy** has emerged from the widened principle of equality of action and reaction or, as we may put it, from the widened principle of conservation of momentum.

The rate at which work is done on a particle is

$$-icF_w,$$

as we have seen **(35·085).** This may be written :

$$-ic\frac{d}{dt}\left(m\frac{dw}{dt}\right) = -ic\frac{d}{dt}(m\,ic)$$

or

$$\frac{d}{dt}(mc^2).$$

Apart from a constant of integration, therefore, the energy of the particle is equal to

$$\boxed{\textbf{mass} \times \mathbf{c^2}.}\ ^{1} \quad . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(35·09)}$$

We must adopt the value zero for the constant of integration in order that the energy of the particle may vanish when it has no mass at all.

The kinetic energy of a particle is

$$mc^2 - m_0c^2, \text{ or } m_0c^2(\gamma - 1) \quad . \quad . \quad . \quad \textbf{(35·091)}$$

since it has to vanish when $v = 0$, i.e. when $\gamma = 1$.

It may easily be verified that for small velocities this reduces, as of course it should, to $m_0v^2/2$. In fact

$$\begin{aligned} \gamma - 1 &= (1 - v^2/c^2)^{-1/2} - 1 \\ &= v^2/2c^2 \end{aligned}$$

for small velocities.

The relationship **(35·09)** makes the principle of conservation of mass a special case of that of conservation of energy. Indeed, we might have defined the measure of energy by

$$\textbf{work done/c}^2,$$

in which case mass would be identical with energy, or perhaps it would be safer to say it would be identical with energy in one of its forms.

The relationship **(35·09)** was derived, for electromagnetic energy, in **§ 27·1,** without apparently invoking relativistic principles. As a matter of fact, however, the derivation there given tacitly assumed that the electromagnetic field equations retain their form under such co-ordinate transformations as we are now considering, which is tantamount to the adoption of the principle of relativity.

Mass, in all the cases where we have encountered it, has been ascribed to something (e.g. a particle or a group of waves) having a definite velocity of translation, whereas energy and momentum are apparently not always associated with something to which such a definite velocity can be assigned, and the concept of energy would therefore seem to be a more comprehensive one than that of mass. It is of course open to us to broaden the definition of **mass** by adopting **(35·09)** for this purpose ; but on the whole it would seem to be desirable—and it is the safe course —to restrict the use of the term to those cases where a quantity of energy is being transported with an assignable velocity (including zero). We may then say that **mass** is one of the forms of **energy.**

[1] Cf. **§ 27·1.**

§ 35·1. Certain Important Tensors

The term ' tensor ' is here used in its general sense and includes scalars or invariants which are tensors of zero rank. The element of ' volume ' in Galilean space-time is an invariant. This can be inferred from the equations of the Lorentz transformation **(34·21)** and **(34·211)** or from those of the widened Lorentz transformation **(34·7)**. In fact

$$\int dx'\,dy'\,dz'\,dw' = \int \begin{vmatrix} \frac{\partial x'}{\partial x}, & \frac{\partial x'}{\partial y}, & \frac{\partial x'}{\partial z}, & \frac{\partial x'}{\partial w} \\ \frac{\partial y'}{\partial x}, & \frac{\partial y'}{\partial y}, & \frac{\partial y'}{\partial z}, & \frac{\partial y'}{\partial w} \\ \frac{\partial z'}{\partial x}, & \frac{\partial z'}{\partial y}, & \frac{\partial z'}{\partial z}, & \frac{\partial z'}{\partial w} \\ \frac{\partial w'}{\partial x}, & \frac{\partial w'}{\partial y}, & \frac{\partial w'}{\partial z}, & \frac{\partial w'}{\partial w} \end{vmatrix} dx\,dy\,dz\,dw \quad . \quad \textbf{(35·1)}$$

The determinant in this equation is identical with the determinant **(34·72)** and therefore equal to + 1, so that

$$\int dx'\,dy'\,dz'\,dw' = \int dx\,dy\,dz\,dw. \quad . \quad . \quad \textbf{(35·101)}$$

The symbols of integration indicate an integration over any region of space-time. The relationship **(35·101)** is indeed to be expected as a consequence of the premiss that Galilean space-time has the same mathematical properties as Euclidean space, extended of course by an extra dimension.

It follows from **(35·101)** that

$$\int dx'\,dy'\,dz'\,dt' = \int dx\,dy\,dz\,dt. \quad . \quad . \quad \textbf{(35·102)}$$

When the spacial part of the integration **(35·101)** is the volume of a body with a rigid boundary, the whole integration may be described as extending over a cylindrical region of space-time; and in a system of reference X_0, Y_0, Z_0, W_0 in which the E velocity of the body vanishes, i.e. one in which the body is at rest relative to the Euclidean rectangular co-ordinates X_0, Y_0, Z_0, the W_0 axis is the axis of the cylindrical region. If now the body is in motion with a constant velocity of translation, v, relative to the axes X, Y, Z, and with a constant velocity of translation, v', relative to the axes X', Y', Z', then

$$dt = \gamma\,d\tau,$$

and

$$dt' = \gamma'\,d\tau,$$

$d\tau$ being the element of proper time and γ and γ' being the functions

$$\gamma = (1 - v^2/c^2)^{-1/2},$$
$$\gamma' = (1 - v'^2/c^2)^{-1/2}.$$

Thus **(35·102)** becomes

$$\gamma' V' \int d\tau = \gamma V \int d\tau$$

where V and V' are respectively the volumes of the body referred to X, Y, Z and that referred to X', Y', Z'. Thus

$$\boxed{\gamma \times \textbf{Euclidean volume of body is invariant}} \quad . \quad . \quad \textbf{(35·11)}$$

This result might of course have been anticipated as a consequence of the FitzGerald-Lorentz contraction.

We have seen (cf. **§ 35·0**) that the Minkowskian force on a body is equal to the Newtonian force multiplied by γ, i.e.

$$\mathscr{F}_x = \gamma F_x,$$
$$\mathscr{F}_y = \gamma F_y,$$

and so on, $\mathscr{F}_x$ being the X component of the Minkowskian force and F_x that of the Newtonian. Now when we divide every component of a vector by the same invariant the result is a vector. If therefore we divide $\mathscr{F}_x$, $\mathscr{F}_y$, $\mathscr{F}_z$, and $\mathscr{F}_w$ by the invariant γV, V being the E volume of the body on which the force is acting, we obtain a Galilean vector and this vector is obviously F_x/V, F_y/V, F_z/V, F_w/V. Thus **the Newtonian force per unit spacial volume is a vector in the Galilean continuum.**

We have agreed, and indeed we are forced, to identify the set of four quantities (s^1, s^2, s^3, s^4) of **(34·011)** with a G vector. This identification is in fact part of the foundation on which our theory is built. Therefore

$$4\pi A \rho v_x / a$$

is the X component of a G vector, and consequently so is

$$\rho \, dx/dt,$$

since the factor, $4\pi A/a$, is a number chosen once for all for all co-ordinate systems. Thus

$$(\rho/\gamma)(dx/d\tau)$$

is the X component of a G vector and consequently

$$\boxed{\rho/\gamma \textit{ is an invariant}} \quad . \quad . \quad . \quad . \quad \textbf{(35·12)}$$

The fact that (s^1, s^2, s^3, s^4) can be identified **(35·12)** with the

product of an invariant and a Galilean velocity suggests that we should regard it as a **contravariant vector.**

Now the expression

$$(\rho/\gamma) \times \gamma V$$

represents the product of two invariants; so that ρV is an invariant. We thus derive the important result that **an electric charge is an invariant under co-ordinate transformations.**

The scalar product

$$\mathscr{F}_x\, dx + \mathscr{F}_y\, dy + \mathscr{F}_z\, dz + \mathscr{F}_w\, dw$$

is an invariant whether the displacement (dx, dy, dz, dw) is that of the particle on which the force $(\mathscr{F}_x, \mathscr{F}_y, \mathscr{F}_z, \mathscr{F}_w)$ is acting or any other displacement. **We must therefore regard the Minkowskian force, and consequently also the Newtonian force per unit spacial volume, as covariant G vectors.**

Turning now to equation **(34·012)** we have identified s^β with a G vector and the foregoing discussion has led us to regard it as a contravariant vector. It therefore follows that we must regard $F^{\beta\alpha}$ as a **contravariant tensor** of rank 2, since its product with the symbolic covariant vector, $\partial/\partial x^\alpha$, yields the contravariant vector s^β. This has already been anticipated by the notation. The covariance of the tensor $F_{\alpha\beta}$ in **(34·02)** will be demonstrated in the next section.

It is usual and convenient to depart from the conventional way of distinguishing covariant from contravariant vectors when we are dealing with the displacement vector

$$(dx^1, dx^2, dx^3, dx^4).$$

In future we shall represent it by

$$(dx_1, dx_2, dx_3, dx_4)$$

or by dx_α, notwithstanding its contravariant character.

CHAPTER III

THE ELECTROMAGNETIC FIELD

§ 35·2. Stress, Momentum, and Energy in the Electromagnetic Field

WE shall start with the expression **(22·62)** for the (Newtonian) force per unit volume, the charge e being replaced by ρ, the charge per unit volume,

$$\mathbf{f} = \rho\left\{\mathcal{E} + \frac{4\pi A}{a}[\mathbf{v}, \mathbf{D}_m]\right\} \quad . \quad . \quad . \quad \textbf{(35·2)}$$

The X component of this force, for example, may be written

$$f_x = \rho\left\{\mathfrak{E}_x + \frac{4\pi A}{a}v_y D_{mz} - \frac{4\pi A}{a}v_z D_{my}\right\},$$

and since $dw = ic\,dt$ we have $dw/dt = v_w = ic$ and consequently

$$v_w = ia/\sqrt{\mu_0 K_0},$$

where μ_0 and K_0 are respectively the permeability and di-electric constant for empty space, therefore

$$f_x = \rho\left\{\frac{-i\sqrt{\mu_0 K_0}}{a}\mathcal{E}_x v_w + \frac{4\pi A}{a}v_y D_{mz} - \frac{4\pi A}{a}v_z D_{my}\right\}$$

or finally

$$f_x = 0 \times \frac{4\pi A\rho}{a}v_x + D_{mz} \times \frac{4\pi A\rho}{a}v_y + (-D_{my}) \times \frac{4\pi A\rho}{a}v_z + \left(\frac{-i\sqrt{\mu_0 K_0}}{4\pi A}\mathcal{E}_x\right) \times \frac{4\pi A\rho}{a}v_w \quad . \quad . \quad \textbf{(35·21)}$$

We shall now replace the subscripts x, y, z, and w by numerals and since f_x, or f_1, is the X component of a covariant vector,[1] while

$$(4\pi A\rho v_x/a,\ 4\pi A\rho v_y/a,\ 4\pi A\rho v_z/a,\ 4\pi A\rho v_w/a)$$

is identical with the contravariant vector (s^1, s^2, s^3, s^4) we may write **(35·21)** in the form:

$$f_1 = F_{11}s^1 + F_{12}s^2 + F_{13}s^3 + F_{14}s^4, \quad . \quad \textbf{(35·211)}$$

[1] Since the product of force and displacement (contravariant vector) is an invariant.

where $F_{11}, F_{12}, F_{13}, F_{14}$ are the components of a **covariant tensor,** and

$$F_{11} = F_{22} = F_{33} = F_{44} = 0,$$

while

$$\left.\begin{aligned} F_{12} &= \frac{\mu H_z}{4\pi A}, \; F_{31} = \frac{\mu H_y}{4\pi A}, \; F_{23} = \frac{\mu H_x}{4\pi A}, \\ F_{41} &= \frac{i\sqrt{\mu_0 K_0}}{4\pi A}\mathcal{E}_x, \; F_{42} = \frac{i\sqrt{\mu_0 K_0}}{4\pi A}\mathcal{E}_y, \; F_{43} = \frac{i\sqrt{\mu_0 K_0}}{4\pi A}\mathcal{E}_z, \end{aligned}\right\} \quad (35{\cdot}22)$$

and of course $F_{\alpha\beta} = -F_{\beta\alpha}$.[1]

The expressions **(35·22)** are identical with those given in **(34·07)** and **(34·071).** If we use the summation convention we may write f_1 **(35·211)** in the form:

$$f_1 = s^\beta F_{1\beta},$$

and any component, f_α, in the form:

$$f_\alpha = s^\beta F_{\alpha\beta}. \quad . \quad . \quad . \quad . \quad . \quad (35{\cdot}221)$$

We now make use of **(34·01)** and substitute for s^β,

$$s^\beta = \frac{\partial F^{\beta\mu}}{\partial x_\mu},$$

thus obtaining

$$f_\alpha = F_{\alpha\beta}\frac{\partial F^{\beta\mu}}{\partial x_\mu}. \quad . \quad . \quad . \quad . \quad (35{\cdot}222)$$

Now we should like, if possible, to express f_α in the form of a divergence, as in **(26·94).** To achieve this we begin by writing **(35·222)** in the following equivalent form:

$$f_\alpha = \frac{\partial}{\partial x_\mu}(F_{\alpha\beta}F^{\beta\mu}) - F^{\beta\mu}\frac{\partial F_{\alpha\beta}}{\partial x_\mu} \quad . \quad . \quad (35{\cdot}23)$$

and make use of equations **(34·02),**

$$\frac{\partial F_{\alpha\beta}}{\partial x_\mu} + \frac{\partial F_{\mu\alpha}}{\partial x_\beta} + \frac{\partial F_{\beta\mu}}{\partial x_\alpha} = 0.$$

Therefore

$$F^{\beta\mu}\frac{\partial F_{\alpha\beta}}{\partial x_\mu} + F^{\beta\mu}\frac{\partial F_{\mu\alpha}}{\partial x_\beta} + F^{\beta\mu}\frac{\partial F_{\beta\mu}}{\partial x_\alpha} = 0.$$

In the middle term the indices β and μ may be interchanged, since it does not matter at all what letter is used to indicate a summation provided the same letter is not being used for some other purpose in a way likely to cause confusion. Thus

$$F^{\beta\mu}\frac{\partial F_{\alpha\beta}}{\partial x_\mu} + F^{\mu\beta}\frac{\partial F_{\beta\alpha}}{\partial x_\mu} + F^{\beta\mu}\frac{\partial F_{\beta\mu}}{\partial x_\alpha} = 0.$$

[1] It will be proved later (**§ 36·6**) that $F_{\alpha\beta}$ can always be given the form: $\partial A_\beta/\partial x_\alpha - \partial A_\alpha/\partial x_\beta$, whatever sort of co-ordinate system we employ, where **A** is a covariant vector.

Interchange of indices in $F^{\mu\beta}$ and $F_{\beta\alpha}$ in the middle term will change the sign of both, so that the term will remain unaltered; and further it is desirable to replace β and μ in the last term by other more non-committal symbols, say σ and τ. Therefore

$$F^{\beta\mu}\frac{\partial F_{\alpha\beta}}{\partial x_\mu} = -\frac{1}{2}F^{\sigma\tau}\frac{\partial F_{\sigma\tau}}{\partial x_\alpha}.$$

If we confine our attention to the special case where μ and K are constants, equations **(35·22)** and **(34·011)** indicate that each $F_{\alpha\beta}$ is equal to the product of $F^{\alpha\beta}$ and a constant, so that the last equation may be written:

$$F^{\beta\mu}\frac{\partial F_{\alpha\beta}}{\partial x_\mu} = -\frac{1}{4}\frac{\partial}{\partial x_\alpha}(F^{\sigma\tau}F_{\sigma\tau}). \quad . \quad . \quad \textbf{(35·24)}$$

We shall now adopt a device which plays a great part in tensor calculations in the theory of relativity. Let $\delta_\alpha{}^\mu$ represent a mixed tensor whose components are:

$$\delta_\alpha{}^\mu = \begin{cases} 1, & \mu = \alpha, \\ 0, & \mu \neq \alpha. \end{cases}$$

It is easy to prove (cf. **§ 35·6**) that its components in any co-ordinate system will be:

$$\delta'_a{}^m = \begin{cases} 1, & m = a, \\ 0, & m \neq a. \end{cases}$$

We may now give **(35·24)** the form:

$$F^{\beta\mu}\frac{\partial F_{\alpha\beta}}{\partial x_\mu} = -\frac{\partial}{\partial x_\mu}\{\tfrac{1}{4}\delta_\alpha{}^\mu F^{\sigma\tau}F_{\sigma\tau}\}.$$

On substituting this in **(35·23)** we obtain

$$f_\alpha = \frac{\partial}{\partial x_\mu}\{F_{\alpha\beta}F^{\beta\mu} + \tfrac{1}{4}\delta_\alpha{}^\mu F^{\sigma\tau}F_{\sigma\tau}\}, \quad . \quad . \quad \textbf{(35·25)}$$

which we may abbreviate by writing

$$f_\alpha = \frac{\partial}{\partial x_\mu}t_\alpha{}^\mu, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(35·251)}$$

where

$$t_\alpha{}^\mu = F_{\alpha\beta}F^{\beta\mu} + \tfrac{1}{4}\delta_\alpha{}^\mu F^{\sigma\tau}F_{\sigma\tau}. \quad . \quad . \quad \textbf{(35·26)}$$

To make it quite clear what **(35·251)** means, we might write down rather more fully the expression for a particular f_α, say f_2 for example:

$$f_2 = \frac{\partial t_2{}^1}{\partial x_1} + \frac{\partial t_2{}^2}{\partial x_2} + \frac{\partial t_2{}^3}{\partial x_3} + \frac{\partial t_2{}^4}{\partial x_4}.$$

The meanings of the **t**'s, if not already clear, will become so when we evaluate them in terms of familiar things. There are four of the equations **(35·251)**. Three of them we have already studied

in **§ 26·9**, and we shall find, as indeed we might anticipate, that the fourth one, namely

$$f_4 = \frac{\partial t_4{}^{\mu}}{\partial x_{\mu}},$$

expresses Poynting's theorem.

If $A_{\alpha}{}^{\mu}$ be any tensor, covariant in α and contravariant in μ, the tensor $\delta_{\mu}{}^{\alpha}A_{\alpha}{}^{\mu}$, where $\delta^{\alpha}{}_{\mu}$ is the tensor used in the description of $t_{\alpha}{}^{\mu}$ in **(35·25)**, is known as **Laue's tensor.** It is a tensor of rank zero, i.e. a scalar. Obviously the tensor

$$\delta_{\mu}{}^{\alpha}t_{\alpha}{}^{\mu} = t_1{}^1 + t_2{}^2 + t_3{}^3 + t_4{}^4 = 0.$$

In order to express the tensor components, $t_{\alpha}{}^{\mu}$, of **(35·26)** in terms of familiar things, we shall first investigate the sum $F^{\sigma\tau}F_{\sigma\tau}$. This, when written at full length, is

$$\begin{aligned} &F^{12}F_{12} + F^{13}F_{13} + F^{14}F_{14} \\ +\, &F^{21}F_{21} + F^{23}F_{23} + F^{24}F_{24} \\ +\, &F^{31}F_{31} + F^{32}F_{32} + F^{34}F_{34} \\ +\, &F^{41}F_{41} + F^{42}F_{42} + F^{43}F_{43}, \end{aligned}$$

since $F_{\alpha\alpha} = F^{\alpha\alpha} = 0$. To carry out the summation we must make use of **(34·011)** and **(35·22)**. We thus get:

$$\begin{aligned} &\frac{\mu}{4\pi A}H_z{}^2 + \frac{\mu}{4\pi A}H_y{}^2 - \frac{K}{4\pi A}\mathcal{E}_x{}^2 \\ +\, &\frac{\mu}{4\pi A}H_z{}^2 + \frac{\mu}{4\pi A}H_x{}^2 - \frac{K}{4\pi A}\mathcal{E}_y{}^2 \\ +\, &\frac{\mu}{4\pi A}H_y{}^2 + \frac{\mu}{4\pi A}H_x{}^2 - \frac{K}{4\pi A}\mathcal{E}_z{}^2 \\ -\, &\frac{K}{4\pi A}\mathcal{E}_x{}^2 - \frac{K}{4\pi A}\mathcal{E}_y{}^2 - \frac{K}{4\pi A}\mathcal{E}_z{}^2, \end{aligned}$$

and therefore

$$F^{\sigma\tau}F_{\sigma\tau} = \frac{1}{4\pi A}\{2\mu\mathbf{H}^2 - 2K\boldsymbol{\mathcal{E}}^2\}. \quad . \quad . \quad \textbf{(35·27)}$$

This expression is only needed for the evaluation of $t_1{}^1$, $t_2{}^2$, $t_3{}^3$, and $t_4{}^4$; since in all the other components $\delta_{\alpha}{}^{\mu}$ is equal to zero.

In the next place, $t_1{}^1$ is equal to the sum $F_{1\beta}F^{\beta 1}$ plus one-fourth of the quantity **(35·27)**. Now

$$F_{1\beta}F^{\beta 1} = -\frac{\mu}{4\pi A}H_z{}^2 - \frac{\mu}{4\pi A}H_y{}^2 + \frac{K}{4\pi A}\mathcal{E}_x{}^2,$$

and consequently

$$t_1{}^1 = \frac{1}{4\pi A}\{K\mathcal{E}_x{}^2 - \tfrac{1}{2}K\boldsymbol{\mathcal{E}}^2 + \mu H_x{}^2 - \tfrac{1}{2}\mu\mathbf{H}^2\}. \quad \textbf{(35·271)}$$

This is identical with the expression for t_{xx} in **(26·95)**. The

expressions for $t_2{}^2$ and $t_3{}^3$ are at once obtained from that for $t_1{}^1$ by replacing the subscript, x, in **(35·271)** by y and z respectively.

In the next place

$$t_4{}^4 = F_{4\beta}F^{\beta 4} + \frac{1}{4\pi A}\left\{\frac{\mu}{2}\mathbf{H}^2 - \frac{K}{2}\mathfrak{E}^2\right\}$$

and therefore

$$t_4{}^4 = \frac{K\mathfrak{E}^2}{8\pi A} + \frac{\mu\mathbf{H}^2}{8\pi A}$$

or

$$t_4{}^4 = U = \begin{pmatrix}\textit{Energy per unit}\\ \textit{spacial volume}\end{pmatrix}. \quad . \quad . \quad \textbf{(35·272)}$$

The reader will have no difficulty in establishing that

$$\left.\begin{aligned} t_2{}^1 = t_1{}^2 &= \frac{1}{4\pi A}\{K\mathcal{E}_x\mathcal{E}_y + \mu H_xH_y\},\\ t_3{}^2 = t_2{}^3 &= \frac{1}{4\pi A}\{K\mathcal{E}_y\mathcal{E}_z + \mu H_yH_z\},\\ t_1{}^3 = t_3{}^1 &= \frac{1}{4\pi A}\{K\mathcal{E}_z\mathcal{E}_x + \mu H_zH_x\},\end{aligned}\right\} \quad . \quad \textbf{(35·273)}$$

and that further

$$t_1{}^4 = \frac{-ic\mu K}{4\pi Aa}[\mathfrak{E},\ \mathbf{H}]_x,$$

or

$$t_1{}^4 = -ic.\frac{\mu K}{a^2}.\frac{a}{4\pi A}[\mathfrak{E},\ \mathbf{H}]_x.$$

This and the two associated equations may therefore be written:

$$\left.\begin{aligned} t_1{}^4 &= -ic\,p_x/u^2,\\ t_2{}^4 &= -ic\,p_y/u^2,\\ t_3{}^4 &= -ic\,p_z/u^2,\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad \textbf{(35·28)}$$

u being the phase velocity of simple harmonic waves in the medium, while $\mathbf{p} \equiv (p_x,\ p_y,\ p_z)$ is Poynting's vector.

The components $t_4{}^1$, $t_4{}^2$, and $t_4{}^3$ turn out, rather disconcertingly, to differ from the three components just given. They are

$$\left.\begin{aligned} t_4{}^1 &= -i\frac{p_x}{c} = \frac{u^2}{c^2}t_1{}^4,\\ t_4{}^2 &= -i\frac{p_y}{c} = \frac{u^2}{c^2}t_2{}^4,\\ t_4{}^3 &= -i\frac{p_z}{c} = \frac{u^2}{c^2}t_3{}^4.\end{aligned}\right\} \quad . \quad . \quad . \quad \textbf{(35·281)}$$

The asymmetry which appears in equations **(35·281)**, i.e. the asymmetry $t_4{}^\alpha \neq t_\alpha{}^4$ ($\alpha = 1, 2, 3$), is rather disturbing when

we remember the symmetry in **(35·273)**, i.e. $t_\beta{}^\alpha = t_\alpha{}^\beta$ (α and $\beta = 1, 2, 3$). Moreover, the symmetry of the tensor **t** of classical Newtonian physics (cf. **§§ 9·7** and **10·5**) leads us to expect symmetry in the present tensor, **t**, since our Galilean space-time and new dynamics are the 4-dimensional analogue of Euclidean space and Newtonian dynamics. The equations we have so far developed undoubtedly apply correctly to a medium which is at rest relatively to the E co-ordinate system, but they lack the property of covariance. In other words they do not retain their form when we pass from one co-ordinate system to another, the reason being that when the medium is at rest in one system it must be in motion in another and our equations have taken no account of this. We shall therefore confine our attention to free space only so that the phase velocity, u, becomes identical with c and the asymmetry of equations **(35·28)** and **(35·281)** disappears. It may be further remarked that we regard material media as constituted in the last analysis of charged particles in otherwise free space so that the ultimate electric fields we have to deal with are in free space and the dielectric constant, K, and permeability, μ, become respectively the invariants K_0 and μ_0 which are characteristic of free space. Occasionally, when it happens to be helpful, we shall use the letter, u, for the phase velocity of electromagnetic waves, even when it is identical with c, to mark the formal difference between it and the c which appears, for example, in $w = ict$.

The last of the equations **(35·25)** we have already suspected to express Poynting's theorem. Let us examine it. It is

$$f_4 = \frac{\partial t_4{}^1}{\partial x_1} + \frac{\partial t_4{}^2}{\partial x_2} + \frac{\partial t_4{}^3}{\partial x_3} + \frac{\partial t_4{}^4}{\partial x_4},$$

or, if we make use of the expressions for $t_4{}^1$, $t_4{}^2$, etc., already given in **(35·281)** and replace $t_4{}^4$ by U the energy per unit volume,

$$f_w = -\frac{i}{c}\,\mathbf{div}\ \mathbf{p} + \frac{\partial U}{\partial w}.$$

We have already learned (cf. **35·085**) that $-icf_4$ or $-icf_w$ represents the rate at which work is done per unit volume by the Newtonian force (f_x, f_y, f_z). It is consequently the rate at which work is done on the charge in the unit volume at the expense of the field energy. Let us therefore multiply both sides of the last equation by $-ic$ and thus obtain

$$-icf_w = -\,\mathbf{div}\ \mathbf{p} - \partial U/\partial t,$$

since $w = ict$; or

$$\boxed{-\partial U/\partial t = \mathbf{div}\ \mathbf{p} - icf_w.} \quad . \quad . \quad . \quad \textbf{(35·29)}$$

This is Poynting's theorem, since $-icf_w$ means the rate at which work is done per unit (spacial) volume at the expense of the electromagnetic energy of the field.

The components t_1^4, t_2^4, and t_3^4 occur in the expressions for f_1, f_2, and f_3 respectively in the terms $\partial t_1^4/\partial x_4$, $\partial t_2^4/\partial x_4$, and $\partial t_3^4/\partial x_4$. The first of these, for example, may be written $\partial(t_1^4/ic)/\partial t$; so that t_1^4/ic is the X component of a momentum reckoned per unit volume. This is the part of the contribution to f_1 which is measured by the rate of increase of t_1^4/ic. The charges or charged particles in the unit volume at a given place gain momentum at the rate $\partial(t_1^4/ic)/\partial t$. The law of conservation of momentum leads us to associate with the field, a *decrease* of momentum per unit volume of the same rate. Consequently we are led to assign to the unit volume of the field a momentum whose X component is

$$-t_1^4/ic + \phi,$$

where ϕ depends only on x, y, and z and will be disregarded for the present. So we conclude (subject to reservations about ϕ) that we may associate with an electromagnetic field a momentum whose X component is $-t_1^4/ic$ per unit volume. Therefore

$$M_x = -t_1^4/ic \quad . \quad . \quad . \quad . \quad \textbf{(35·291)}$$

or

$$\left.\begin{aligned} M_x &= p_x/u^2, \\ M_y &= p_y/u^2, \\ M_z &= p_z/u^2, \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad \textbf{(35·292)}$$

where $\mathbf{p}$ is Poynting's vector and u is the phase velocity of plane harmonic electromagnetic waves in the medium. This result is identical with that found in **§ 26·9.**

§ 35·3. An Analogue of Poynting's Theorem

We learned in **§ 10·2** that the components of the body force, $\mathbf{R}$ (i.e. the force of external origin as distinguished from that due to elastic stress), per unit volume, are expressed by

$$R_x = -\left(\frac{\partial t_{xx}}{\partial x} + \frac{\partial t_{xy}}{\partial y} + \frac{\partial t_{xz}}{\partial z}\right) + \rho\frac{\partial^2\alpha}{\partial t^2}, \quad . \quad \textbf{(35·3)}$$

together with similar equations for the Y and Z components. The t_{xx}, t_{xy}, etc., are the components at some point (X, Y, Z) of the elastic stress tensor which has been defined in **§ 9·7**; (α, β, γ) are the components of the displacement of an element of the medium, situated at that point, from the position it would occupy when undisturbed by forces, while ρ represents the density of the medium at (x, y, z). Finally (R_x, R_y, R_z) represents the

body force per unit volume at the same point. This equation is dealt with in **§ 10·5** for the special case where **R** vanishes, and reference to equation **(10·54)** and its mode of derivation will make it clear that we may write for **(35·3)**, in the special case where the medium consists of particles which have no field of their own and only influence one another by collisions:

$$R_x=\frac{\partial}{\partial x}(\Sigma m v_x^2)+\frac{\partial}{\partial y}(\Sigma m v_x v_y)+\frac{\partial}{dz}(\Sigma m v_x v_z)+\frac{\partial}{\partial t}(\Sigma m v_x). \qquad \textbf{(35·31)}$$

The only differences between this equation and **(10·54)** are (1) that we are no longer confining our attention to the case where **R** vanishes and (2) we now represent the velocity components (u, v, w) of **§ 10·5** by (v_x, v_y, v_z). The letter m represents the mass of any particle, (v_x, v_y, v_z) its velocity, and the summation extends over all the particles in the unit volume; or, to be more precise, it represents the summation over the particles in a volume element $dx\,dy\,dz$, the result being divided by $dx\,dy\,dz$.

On multiplying and dividing the last term of **(35·31)** by ic (where c is the velocity of electromagnetic waves in empty space) we get, since $ic = v_w$ and $ict = w$,

$$R_x = \frac{\partial}{\partial x}(\Sigma m v_x^2) + \frac{\partial}{\partial y}(\Sigma m v_x v_y) + \frac{\partial}{\partial z}(\Sigma m v_x v_z) + \frac{\partial}{\partial w}(\Sigma m v_x v_w). \quad . \quad . \quad \textbf{(35·311)}$$

There are, of course, similar equations for R_y and R_z and a fourth equation is suggested, namely:

$$R_w = \frac{\partial}{\partial x}(\Sigma m v_w v_x) + \frac{\partial}{\partial y}(\Sigma m v_w v_y) + \frac{\partial}{\partial z}(\Sigma m v_w v_z) + \frac{\partial}{dw}(\Sigma m v_w^2). \quad . \quad . \quad \textbf{(35·32)}$$

In the special case where the components of **R** vanish these equations are identical with **(10·55)**, except for the slight difference in notation, and the fact that there c was an entirely undetermined velocity, not identified with the velocity of light, or of electromagnetic waves, in empty space. We now know that it should be identified with the product of the velocity of electromagnetic waves in empty space and $\sqrt{-1}$.

It should be carefully borne in mind that equations **(35·31)** et seq. refer expressly to a medium constituted of particles, such as those studied in **§ 35**, which influence one another by impact only. If we wish to leave the question of the constitution of the medium, whether a material medium or an electromagnetic field

or whatever it may be, quite open, we should give equations **(35·311)** and **(35·32)** the form:

$$\left.\begin{aligned} R_1 &= \frac{\partial p_1{}^1}{\partial x_1} + \frac{\partial p_1{}^2}{\partial x_2} + \frac{\partial p_1{}^3}{\partial x_3} + \frac{\partial p_1{}^4}{\partial x_4}, \\ R_4 &= \frac{\partial p_4{}^1}{\partial x_1} + \frac{\partial p_4{}^2}{\partial x_2} + \frac{\partial p_4{}^3}{\partial x_3} + \frac{\partial p_4{}^4}{\partial x_4}, \end{aligned}\right\} \quad . \quad . \quad \textbf{(35·33)}$$

or, alternatively, the form:

$$\left.\begin{aligned} f_1 &= \frac{\partial t_1{}^1}{\partial x_1} + \frac{\partial t_1{}^2}{\partial x_2} + \frac{\partial t_1{}^3}{\partial x_3} + \frac{\partial t_1{}^4}{\partial x_4}, \\ f_4 &= \frac{\partial t_4{}^1}{\partial x_1} + \frac{\partial t_4{}^2}{\partial x_2} + \frac{\partial t_4{}^3}{\partial x_3} + \frac{\partial t_4{}^4}{\partial x_4}. \end{aligned}\right\} \quad . \quad . \quad \textbf{(35·331)}$$

In these equations **R** means the force (per unit volume) impressed from outside. It might, for example, be due to gravity. On the other hand, **f** means the force per unit volume due to the state of stress evoked in the medium, or to the field; so that

$$\mathbf{f} = -\mathbf{R}.$$

This is simply an expression of the law of action and reaction with which the laws of conservation of mass, momentum, and energy are associated. Obviously $t_1{}^1$, $t_1{}^2$, etc., are equal to $-p_1{}^1$, $-p_1{}^2$, etc., respectively. The components $t_1{}^1$, $t_2{}^2$, $t_3{}^3$ are of the nature of tensions, while $p_1{}^1$, $p_2{}^2$, $p_3{}^3$ are of the nature of pressures (cf. **§§ 9·7** and **9·9**).

It is instructive to examine the equation **(35·32)** more closely. The left-hand member, namely R_w, if multiplied by $-ic$, represents the rate at which the **external** or **impressed force** per unit volume is doing work (cf. **35·051**). Multiply both sides of the equation by $-ic$ and remember that $v_w = ic$. We thus get

$$-icR_w = \frac{\partial}{\partial x}(\Sigma mc^2 v_x) + \frac{\partial}{\partial y}(\Sigma mc^2 v_y) + \frac{\partial}{\partial z}(\Sigma mc^2 v_z) + \frac{\partial}{\partial t}(\Sigma mc^2),$$

and therefore

$$-icR_w = \mathbf{div}\{(\Sigma mc^2 v_x), (\Sigma mc^2 v_y), (\Sigma mc^2 v_z)\} + \frac{\partial}{\partial t}(\Sigma mc^2),$$

or, if we represent the E vector

$$\{(\Sigma mc^2 v_x), (\Sigma mc^2 v_y), (\Sigma mc^2 v_z)\}$$

by $\mathbf{p}' \equiv (p_x', p_y', p_z')$ and Σmc^2 by U, we get

$$\frac{\partial U}{\partial t} = -\mathbf{div}\ \mathbf{p}' - icR_4. \quad . \quad . \quad . \quad \textbf{(35·34)}$$

This is the exact analogue of Poynting's theorem. The left-hand number means the rate of increase, at some point (x, y, z), of the energy of the medium per unit volume. **div** $\mathbf{p}'$ means the rate,

reckoned per unit volume, at which energy is leaving any small region in the neighbourhood of (x, y, z) through the surface enclosing it; so that $-$ **div p′** means the rate at which it is entering. And lastly we have already learned that $-icR_w$ means the rate at which the external force does work per unit volume of the medium.

Equation **(35·32),** it will be noticed, becomes identical with **(10·55)** when the body force vanishes. In this case it states that the rate of increase of *mass* within the region enclosed by some surface is equal to the rate at which *mass* flows inwards through the surface. Or we might say that in this case the Poynting equation and the equation of continuity **(10·52)** become identical. And we realize that the reason for it is the identity, apart from the trivial difference occasioned by the factor, c^2, of mass and energy.

The complete symmetry of the tensor, **p,** in equations **(35·311)** and **(35·32)** should be noted and also the fact that the vector, **p′,** —the analogue of Poynting's vector—in equation **(35·34)** *really is identical with the energy transport vector*, since it is equal to Σmc^2 **v,** mc^2 being the energy of a single particle and the summation being extended over the unit volume (cf. **§§ 35·2** and **35·5**).

§ 35·4. The Mass of a Charged Spherical Conductor

As an exercise on the theory of stress and momentum, more especially in an electromagnetic field, and because of its importance on its own account, we shall turn back to the problem of **§§ 26·1** and **28** which we can now deal with much more thoroughly

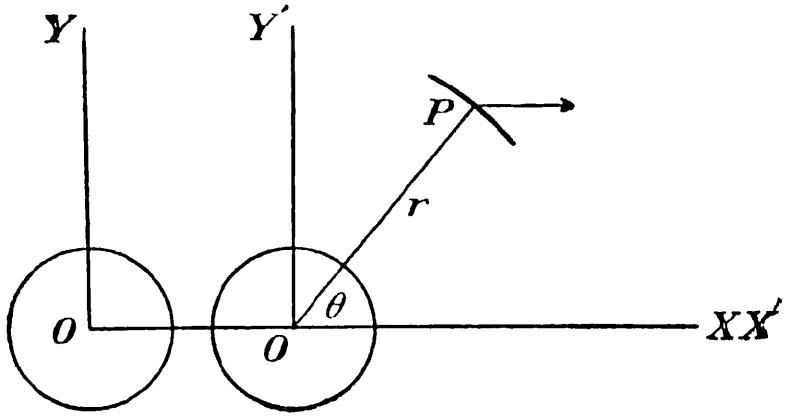

Fig. 35·4

and without the need of *ad hoc* hypotheses. It is convenient to think of the sphere as a conducting liquid—mercury, for example. It is spherical in shape in a co-ordinate system X', Y', Z' (E) in which it is at rest, and we may suppose its centre to coincide with

the origin of this system (Fig. 35·4). It has a velocity, v, along the X axis of a system X, Y, Z (E). The two systems of co-ordinates coincide at the instant $t = t' = 0$. The G equations of transformation (Lorentz transformation) are

$$\left.\begin{aligned} x_1 &= \cos\phi.x_1' + 0 + 0 + \sin\phi.x_4' \\ x_2 &= \quad 0 \quad + x_2' + 0 + 0 \\ x_3 &= \quad 0 \quad + 0 + x_3' + 0 \\ x_4 &= (-\sin\phi).x_1' + 0 + 0 + \cos\phi.x_4', \end{aligned}\right\} \quad . \quad (35\cdot4)$$

or the equivalent equations

$$\left.\begin{aligned} x_1' &= \cos\phi.x_1 + 0 + 0 - \sin\phi.x_4 \\ x_2' &= \quad 0 + x_2 + 0 + 0 \\ x_3' &= \quad 0 + 0 + x_3 = 0 \\ x_4' &= \sin\phi.x_1 + 0 + 0 + \cos\phi.x_4. \end{aligned}\right\} \quad . \quad (35\cdot401)$$

These are identical with the equations **(34·21)**, **(34·211)** and **(34·23)** except for the trivial difference that we are now using numerical subscripts to distinguish the co-ordinates.

Our method will consist in calculating the X component of the momentum per unit volume at a point in the field outside the sphere, integrating over the whole region outside the sphere and dividing the result by v. We begin, therefore, by finding a suitable expression for t_1^4. Now

$$t_1^4 = \frac{\partial x_4}{\partial x_\mu'} \frac{\partial x_\nu'}{\partial x_1} t'^{\mu}_{\nu},$$

(vide **§ 34·8**), thus

$$t_1^4 = -\sin\phi.\frac{\partial x_\nu'}{\partial x_1} t'^{1}_{\nu} + \cos\phi.\frac{\partial x_\nu'}{\partial x_1} t'^{4}_{\nu},$$

when we sum with respect to μ and use equations **(35·4)**. On summing with respect to ν and using equations **(35·401)** we get

$$t_1^4 = -\sin\phi.\cos\phi.t'^{1}_{1} - \sin^2\phi.t'^{1}_{4} + \cos^2\phi.t'^{4}_{1} + \cos\phi.\sin\phi.t'^{4}_{4}.$$

The components t'^{1}_{4} and t'^{4}_{1} vanish, since the sphere is at rest in X', Y', Z'. Therefore

$$t_1^4 = \cos\phi.\ \sin\phi(t'^{4}_{4} - t'^{1}_{1}). \quad . \quad . \quad . \quad (35\cdot41)$$

Now t'^{4}_{4} is the energy per unit volume in the system X', Y', Z'. Therefore

$$t'^{4}_{4} = \frac{K_0}{8\pi A}\mathfrak{E}'^2$$

and

$$t'^{1}_{1} = \frac{K_0}{4\pi A}\mathfrak{E}'^2_x - \frac{K_0}{8\pi A}\mathfrak{E}'^2,$$

while M_x, the X component of the momentum per unit volume, is equal to $-t_1^4/ic$ and

$$\cos\phi\sin\phi = \gamma^2 v/ic.$$

On substituting these values in **(35·41)** we easily get

$$M_x = \frac{\gamma^2 v}{c^2} \cdot \frac{K_0}{4\pi A}(\mathfrak{E}'^2 - \mathfrak{E}'_x{}^2), \quad . \quad . \quad . \quad \textbf{(35·42)}$$

and therefore, if θ be the angle between $\mathfrak{E}'$ and the X' axis,

$$M_x = \frac{\gamma^2 v}{c^2} \cdot \frac{K_0}{4\pi A} \mathfrak{E}'^2 \sin^2 \theta. \quad . \quad . \quad . \quad \textbf{(35·43)}$$

Now

$$\mathfrak{E}'^2 = \frac{A^2 e^2}{K_0^2 r^4},$$

where r is the distance, in the X', Y', Z' system, from the centre of the sphere to the point in question. Hence

$$M_x = \frac{\gamma^2 v A e^2}{4\pi c^2 K_0 r^4} \sin^2 \theta \quad . \quad . \quad . \quad \textbf{(35·44)}$$

If we multiply this by the E volume element, $dx\,dy\,dz$, we get the X component of the momentum in it; and, since

$$\gamma\, dx\, dy\, dz = dx'\, dy'\, dz',$$

therefore

$$M_x\, dx\, dy\, dz = \frac{\gamma v A e^2 \sin^2 \theta}{4\pi c^2 K_0 r^4} dx'\, dy'\, dz'.$$

Now let us replace $dx'\, dy'\, dz'$ in this expression by $r^2\, dr \sin\theta\, d\theta\, d\phi$ and integrate over the whole field, so that the total X component of the momentum—which is the total electromagnetic momentum of the charged sphere, since the remaining components obviously vanish—becomes

$$\frac{\gamma v A e^2}{4\pi c^2 K_0} \int_R^\infty \int_0^\pi \int_0^{2\pi} \frac{\sin^3 \theta\, dr\, d\theta\, d\phi}{r^2},$$

where R is the radius of the resting sphere. The integral is equal to $8\pi/3R$ and so we get

$$\boxed{\text{Electromagnetic momentum} = \frac{2\gamma A e^2}{3K_0 R c^2} \cdot v} \quad . \quad \textbf{(35·45)}$$

and

$$\boxed{\text{Electromagnetic mass} = \frac{2\gamma A e^2}{3K_0 R c^2}} \quad . \quad . \quad \textbf{(35·46)}$$

When the velocity, **v**, of the sphere is very small, $\gamma = 1$ and so the electromagnetic 'rest' mass is

$$\frac{2Ae^2}{3K_0 R c^2}$$

in complete agreement with the results of the calculations in **§§ 26·1** and **28**, since $a^2/K_0\mu_0 = c^2$.

Let us now investigate the state of affairs within the spherical surface. We have thought of the spherical conductor as a liquid. It is therefore subject, when regarded from the X', Y', Z' (E) system of co-ordinates in which it is at rest, to a uniform tension in consequence of the charge distributed over its surface. This tension is equal to $K_0\mathfrak{E}'^2/8\pi A$, where $\mathfrak{E}'$ is the field intensity at the surface of the sphere. The uniform tension within the spherical boundary is therefore equal to $Ae^2/8\pi K_0 R^4$; or we may regard it as subject to a uniform pressure, p', where

$$p' = -Ae^2/8\pi K_0 R^4. \quad . \quad . \quad . \quad . \quad \textbf{(35·47)}$$

In addition to this we have to remember that there is a certain amount of energy, U', per unit volume, part of it contributed by the mass density of the material of the sphere in its unstrained condition, the rest being the strain energy per unit volume. We may represent the 16 components of the stress momentum tensor within the sphere and referred to the G system of reference (X_1, X_2, X_3, X_4) by $p_1{}^1, p_1{}^2, \ldots p_2{}^1, p_2{}^2, \ldots p_3{}^1, p_3{}^2, \ldots$ etc. In the system (X_1', X_2', X_3', X_4') these reduce to

$$p'_1{}^1 = p'_2{}^2 = p'_3{}^3 = -Ae^2/8\pi K_0 R^4$$

(vide **§ 19·2,** more especially the footnote on page 44 of Vol. II)

and $$p'_4{}^4 = -U'.$$

Now the X component of the momentum per unit volume is $p_1{}^4/ic$ and

$$p_1{}^4 = \cos\phi \sin\phi(p'_4{}^4 - p'_1{}^1)$$

(cf. **35·41**). The X component of the momentum per unit volume is consequently

$$\frac{\gamma^2 v}{c^2}\left(U' - \frac{Ae^2}{8\pi K_0 R^4}\right).$$

This is constant throughout the volume of the conductor and the total internal momentum is therefore obtained by multiplying it by the volume of the conductor, namely, $4\pi R^3/3\gamma$, since the volume of the sphere at rest is $4\pi R^3/3$. We thus get for the momentum within the conductor

$$\frac{\gamma v}{c^2}\left\{U' - \frac{Ae^2}{8\pi K_0 R^4}\right\}\frac{4\pi R^3}{3},$$

or

$$\gamma\left\{\frac{E'}{c^2} - \frac{Ae^2}{6K_0 Rc^2}\right\}v, \quad . \quad . \quad . \quad . \quad \textbf{(35·48)}$$

where E' means the energy within the conductor, referred to the

system of reference within which it is at rest. Therefore the mass associated with the interior of the sphere is

$$\gamma\left\{\frac{E'}{c^2}-\frac{Ae^2}{6K_0Rc^2}\right\}. \quad . \quad . \quad . \quad . \quad . \quad (35{\cdot}481)$$

The total mass, it will be seen, is made up of the two contributions **(35·46)** and **(35·481)** and is therefore equal to

$$\gamma\left\{\frac{E'}{c^2}-\frac{Ae^2}{6K_0Rc^2}+\frac{2Ae^2}{3K_0Rc^2}\right\},$$

or

$$\gamma\left\{E'+\frac{Ae^2}{2K_0R}\right\}/c^2,$$

or

$$\left\{\begin{matrix}\textbf{Total mass of}\\ \textbf{moving sphere}\end{matrix}\right\}=\gamma\times\left\{\begin{matrix}\textbf{Total energy of}\\ \textbf{sphere at rest}\end{matrix}\right\}/c^2 \quad . \quad (35{\cdot}49)$$

§ 35·5. Einstein's Mass-Energy Formula

One of the conclusions we reached in studying the dynamics of the special theory of relativity **(§ 35)**—and still earlier **(§ 27·1)** for the special case of an electromagnetic wave group—is that the mass of a particle or moving system is equal to its energy divided by c^2. We also found that the mass of a particle is equal to the product of γ and the rest mass. These results are confirmed in **(35·49)**; but it should be observed that the mass-energy relation may fail when we apply it to parts of a system which are incapable of a separate existence. For example, it does not hold for the electromagnetic field of the sphere by itself, since as we have seen its mass as given by **(35·46)** is $2\gamma Ae^2/3K_0Rc^2$ and the electromagnetic energy of the sphere at rest is $Ae^2/2K_0R$. Consequently the electromagnetic mass as given by **(35·46)** is

$$\tfrac{4}{3}\gamma\{\textbf{Electromagnetic energy at rest}\}/c^2,$$

that is 4/3 of that given by Einstein's law.

This will not surprise us when we remember that Poynting's vector, which is proportional to the three components $t_4{}^1$, $t_4{}^2$, and $t_4{}^3$ in **(35·331)** cannot, except in special cases, be identified with the energy transport vector. Indeed, its components may be replaced by those of any other vector possessing the same divergence without interfering with the validity of Poynting's theorem. We notice too that the momentum (per unit volume) when we take its components to be proportional to $t_1{}^4$, $t_2{}^4$, and $t_3{}^4$, has in general a different direction from that in which the

electromagnetic field is convected and vanishes at points on the axis where the energy density is by no means zero.

§ 35·6. The Electromagnetic Vector Potential

In what follows μ and K have the same values as in empty space. The equations **(34·02)** can be satisfied by writing:

$$F_{\alpha\beta} = \frac{\partial A_\beta}{\partial x_\alpha} - \frac{\partial A_\alpha}{\partial x_\beta} \quad . \quad . \quad . \quad . \quad . \quad \textbf{(35·6)}$$

This statement does not of course completely define the vector **A**, since it remains true after we superpose on **A** any vector which is a gradient, i.e. any vector

$$\left(\frac{\partial \phi}{\partial x_1}, \frac{\partial \phi}{\partial x_2}, \frac{\partial \phi}{\partial x_3}, \frac{\partial \phi}{\partial x_4}\right), \quad . \quad . \quad . \quad \textbf{(35·601)}$$

where ϕ is any scalar, i.e. invariant, quantity. If we represent the symbolic vector

$$\textbf{grad} \equiv \left(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3}, \frac{\partial}{\partial x_4}\right)$$

by

$$\square \equiv (\square_1, \square_2, \square_3, \square_4)$$

we may represent the gradient **(35·601)** by

$$\square\phi. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(35·602)}$$

Now the relationship between the covariant component $F_{\alpha\beta}$, and the corresponding contravariant one, $F^{\alpha\beta}$, is

$$F_{\alpha\beta} = \frac{\mu}{4\pi A} F^{\alpha\beta} \quad . \quad . \quad . \quad . \quad . \quad \textbf{(35·61)}$$

(cf. equations **34·03, 34·04,** and **34·061**). This equation **(35·61)**, though quantitatively correct, is formally in conflict with the distinction we have made between covariant and contravariant tensors. In order to preserve the formal distinction we shall introduce certain new tensors. If $g_n{}^m$ be a mixed tensor of the second rank and if its components have the values

$$g_n{}^m = \begin{cases} 1, & m = n, \\ 0, & m \neq n, \end{cases} \quad . \quad . \quad . \quad . \quad \textbf{(35·62)}$$

in the (X_1, X_2, X_3, X_4) co-ordinate system, they will have the values

$$g'_\nu{}^\mu = \begin{cases} 1, & \mu = \nu, \\ 0, & \mu \neq \nu, \end{cases} \quad . \quad . \quad . \quad . \quad \textbf{(35·621)}$$

in any other system (X_1', X_2', X_3', X_4') the proof is simple. Since $g_n{}^m$ is a mixed tensor

$$g'_\nu{}^\mu = \frac{\partial x'_\mu}{\partial x_m} \frac{\partial x_n}{\partial x'_\nu} g_n{}^m,$$

we get on carrying out the summation with respect to m and n

$$g'_{\nu}{}^{\mu} = \frac{\partial x_\mu'}{\partial x_\nu'},$$

by **(35·62)** and since $\partial x_\mu'/dx_\nu' = 1$ or 0 according as $\mu = \nu$ or $\mu \neq \nu$, the result **(35·621)** must follow. We can see now that we may, when convenient, adopt two further tensors, g_{mn} and g^{mn}, one covariant and the other contravariant with the property **(35·62).** We can be confident of this since the distinction between covariance and contravariance is purely formal, so long as we are using orthogonal co-ordinate systems.

We may now give **(35·61)** the form:

$$g^{\alpha m}g^{\beta n}F_{mn} = \frac{\mu}{4\pi A}F^{\alpha\beta}, \quad . \quad . \quad . \quad . \quad \textbf{(35·63)}$$

which is both quantitatively and formally correct. The symbol μ represents, it will be remembered, the permeability of empty space and we shall no longer distinguish it by writing μ_0.

Now

$$\frac{\partial F^{\alpha\beta}}{\partial x_\beta} = s^\alpha,$$

(cf. **34·01** and **34·012**). Therefore

$$\frac{4\pi A}{\mu}\ \frac{\partial}{\partial x_\beta}\{g^{\alpha m}g^{\beta n}F_{mn}\} = s^\alpha = \frac{4\pi A}{a}\rho v^\alpha,$$

or

$$\frac{\partial}{\partial x_\beta}\left\{g^{\alpha m}g^{\beta n}\left(\frac{\partial A_n}{\partial x_m} - \frac{\partial A_m}{\partial x_n}\right)\right\} = \sigma^\alpha, \quad . \quad . \quad \textbf{(35·64)}$$

where

$$\sigma^\alpha = \mu\rho v^\alpha/a. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(35·641)}$$

Multiply **(35·64)** by $g_{\alpha\gamma}$ on both sides and carry out the indicated summations, thus:

$$\frac{\partial}{\partial x_\gamma}\left(\frac{\partial A_\beta}{\partial x_\beta}\right) = \frac{\partial}{\partial x_\beta}\left(g^{\beta n}\frac{\partial A_\gamma}{\partial x_n}\right) + \sigma_\gamma,$$

or

$$\frac{\partial}{\partial x_\gamma}\left\{\frac{\partial A_1}{\partial x_1} + \frac{\partial A_2}{\partial x_2} + \frac{\partial A_3}{\partial x_3} + \frac{\partial A_4}{\partial x_4}\right\}$$

$$= \left\{\frac{\partial^2}{\partial x_1{}^2} + \frac{\partial^2}{\partial x_2{}^2} + \frac{\partial^2}{\partial x_3{}^2} + \frac{\partial^2}{\partial x_4{}^2}\right\}A_\gamma + \sigma_\gamma,$$

which may be written shortly

$$\square_\gamma((\square \mathrm{A})) = \square^2 A_\gamma + \sigma_\gamma. \quad . \quad . \quad . \quad \textbf{(35·65)}$$

Let us suppose that we have found a solution, A′, of **(35·65)**; then since **(35·6)** is not affected by superposing on A′ any vector

which is a gradient, **(35·65)** will be satisfied when we substitute for **A** the vector

$$\mathbf{A}'' = \mathbf{A}' + \Box\phi,$$

ϕ being any scalar quantity. On making the substitution we get

$$\Box_\gamma\{((\Box,\ \mathbf{A}')) + \Box^2\phi\} = \Box^2\mathbf{A}_\gamma'' + \sigma_\gamma.$$

The quantity in the brackets, { }, on the left, it will be observed, is common to all four equations, $\gamma = 1, 2, 3, 4$. $\mathbf{A}'$ is *any* solution of **(35·65)** and is given, by hypothesis; while ϕ may be any scalar quantity whatever. Let us choose ϕ so that

$$\Box^2\phi = -((\Box,\ \mathbf{A}')),$$

the solution, $\mathbf{A}''$, will then be determined by the equations

$$\Box^2\mathbf{A}_\gamma'' = -\sigma_\gamma, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(35·66)}$$

any of which may be called a 4-dimensional or extended **Poisson's equation** (cf. **§§ 27·4** and **27·5** in the latter of which an important solution of the extended Poisson's equation is given).

CHAPTER IV

INTRODUCTION TO THE GENERAL THEORY OF RELATIVITY

§ 35·7. Motion of a Free Particle

NEWTONIAN dynamical theory and the special theory of relativity, both of them, prescribe a rectilinear path for the motion of a particle not subject to outside influences. Its equations of motion are:

$$\left.\begin{aligned} \frac{d^2x_\alpha}{d\tau^2} &= 0, \\ \text{or, if preferred,} \qquad \\ \frac{d^2x_\alpha}{ds^2} &= 0. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad (35{\cdot}7)$$

Now it will be remembered that Hamilton was able to describe the motion of a large class of systems **(§ 8·6)** by the single statement:

$$\delta\int_1^2 (T - V)dt = 0,$$

or

$$\delta\int_1^2 (2T - E)dt = 0,$$

in which there is no explicit reference to any system of co-ordinates. The integral (Hamilton's principal function) to which a stationary value is assigned is an invariant, and its invariant analogue in the special theory of relativity is

$$\int_1^2 (p_x\,dx + p_y\,dy + p_z\,dz - E\,dt).$$

In the case of a single particle this may be written:

$$\left.\begin{aligned} &\int_1^2 (p_x\,dx + p_y\,dy + p_z\,dz + p_w\,dw), \\ \text{or} \qquad &\int_1^2 ((\mathbf{p},\ \mathbf{ds})), \end{aligned}\right\} \quad . \quad (35{\cdot}71)$$

since it is the integral of the 4-dimensional scalar product of

$$\mathbf{p} \equiv (p_x,\ p_y,\ p_z,\ p_w)$$

and

$$\mathbf{ds} \equiv (dx,\ dy,\ dz,\ dw).$$

Applying the stationary principle to **(35·71)** we have, since

$$p_x = m_0 \frac{dx}{d\tau},$$

$$p_y = m_0 \frac{dy}{d\tau},$$

. . . . .

$$p_w = m_0 \frac{dw}{d\tau},$$

$$\delta \int_1^2 m_0 \left\{ \left(\frac{dx}{d\tau}\right)^2 + \left(\frac{dy}{d\tau}\right)^2 + \left(\frac{dz}{d\tau}\right)^2 + \left(\frac{dw}{d\tau}\right) \right\}^2 d\tau = 0,$$

and therefore, since

$$dx^2 + dy^2 + dz^2 + dw^2 = -\, c^2\, d\tau^2$$

(cf. **34·9**),

$$\delta \int_1^2 (-\, m_0 c^2) d\tau = 0,$$

or

$$\delta \int_1^2 d\tau = 0.$$

This is equivalent of course to

$$\delta \int_1^2 ds = 0. \qquad . \ . \ . \ . \ . \quad \textbf{(35·72)}$$

The length of the path, measured by the sum of the elementary intervals, ds, which make it up, has a stationary value (cf. **§ 28·4**).

We may carry out the variation in the following way. We have

$$\int_1^2 \delta \{ds^2\}^{1/2} = 0,$$

or

$$\int_1^2 \frac{1}{2\, ds} \delta \{ds^2\} = 0,$$

and therefore

$$\int_1^2 \frac{1}{2\, ds} \{2\, dx\, \delta\, dx + 2\, dy\, \delta\, dy + 2\, dz\, \delta\, dz + 2\, dw\, \delta\, dw\} = 0,$$

or

$$\int_1^2 \frac{1}{ds} \{dx\, d\, \delta x + dy\, d\, \delta y + dz\, d\, \delta z + dw\, d\, \delta w\} = 0,$$

since we may interchange d and δ.[1] Therefore

$$\int_1^2 \frac{1}{ds}\{d(dx\ \delta x) - \delta x\ d^2x + d(dy\ \delta y) - \delta y\ d^2y + d(dz\ \delta z) - \delta z\ d^2z + d(dw\ \delta w) - \delta w\ d^2w\} = 0. \quad . \quad (35{\cdot}721)$$

Now each successive ds in the integral **(35·721)** is a ds along the actual path of the particle and it is clearly permissible to regard these successive intervals, ds, as all equal if we wish. The positive part of **(35·721)** may therefore, if we choose, be written:

$$\frac{1}{ds}\int_1^2 d\{dx\ \delta x + dy\ \delta y + dz\ \delta z + dw\ \delta w\}.$$

This obviously vanishes, since the variations δx, δy, etc., vanish at the limits, 1 and 2, of the integral. Hence **(35·721)** becomes

$$\int_1^2 ds\left\{\frac{d^2x}{ds^2}\delta x + \frac{d^2y}{ds^2}\delta y + \frac{d^2z}{ds^2}\delta z + \frac{d^2w}{ds^2}\delta w\right\} = 0.$$

Hence

$$\frac{d^2x}{ds^2} = \frac{d^2y}{ds^2} = \frac{d^2z}{ds^2} = \frac{d^2w}{ds^2} = 0$$

in agreement with **(35·7)**.

We may of course make use of co-ordinates which are not orthogonal; polar co-ordinates, for example, in which case the expression for the square of the interval will be

$$ds^2 = dr^2 + r^2 d\,\theta^2 + r^2 \sin^2\phi\, d\phi^2 - c^2\, dt^2. \quad . \quad (35{\cdot}73)$$

The equations of motion of a free particle, in these co-ordinates, are easily found to be

$$\left.\begin{aligned} &\frac{d^2r}{d\tau^2} - r\left(\frac{d\theta}{d\tau}\right)^2 - r\sin^2\theta\left(\frac{d\phi}{d\tau}\right)^2 = 0,\\ &\frac{d}{d\tau}\left(r^2\frac{d\theta}{d\tau}\right) - r^2\sin\theta\cos\theta\left(\frac{d\phi}{d\tau}\right)^2 = 0,\\ &\frac{d}{d\tau}\left(r^2\sin^2\theta\frac{d\phi}{d\tau}\right) = 0,\\ &\frac{d^2t}{d\tau^2} = 0, \end{aligned}\right\} \quad . \quad (35{\cdot}74)$$

whether we start out from **(35·7)** or **(35·72)** or from Hamilton's canonical equations **(8·43** and **8·46)**.

[1] Cf. **§ 3·3.**

More generally we may replace the orthogonal co-ordinates by new co-ordinates defined by arbitrarily chosen functions of them. Suppose our orthogonal co-ordinates are x_1', x_2', x_3', and x_4', in which case the interval is expressed by

$$ds^2 = dx'_1{}^2 + dx'_2{}^2 + dx'_3{}^2 + dx'_4{}^2. \quad . \quad . \quad \textbf{(35·75)}$$

If we introduce new co-ordinates x_1, x_2, x_3, x_4, defined by

$$\begin{aligned} x_1' &= f_1(x_1, x_2, x_3, x_4), \\ x_2' &= f_2(x_1, x_2, x_3, x_4), \\ x_3' &= f_3(x_1, x_2, x_3, x_4), \\ x_4' &= f_4(x_1, x_2, x_3, x_4), \end{aligned}$$

the interval will now take the form:

$$ds^2 = g_{mn}\, dx_m\, dx_n, \quad . \quad . \quad . \quad . \quad \textbf{(35·76)}$$

the summation convention being understood to be in operation and the coefficients, g_{mn}, being given by

$$g_{mn} = \sum_{\mu} \frac{\partial x_\mu'}{\partial x_m} \frac{\partial x_\mu'}{\partial x_n}.$$

The coefficients, g_{mn}, it will be noticed, are symmetrical, i.e. $g_{mn} = g_{nm}$. The motion of the free particle is of course still expressed by **(35·72)** which is quite independent of our choice of co-ordinates. To find the equations of motion referred to the system of co-ordinates we have chosen, we have to start, as before, from

$$\delta \int_1^2 ds = \delta \int_1^2 \{ds^2\}^{1/2} = \int_1^2 \frac{1}{2\,ds} \delta\{ds^2\} = 0.$$

Therefore

$$\int_1^2 \frac{1}{2\,ds} \delta\{g_{mn}\, dx_m\, dx_n\} = 0,$$

and consequently

$$\int_1^2 \frac{1}{2\,ds} \left\{ \frac{\partial g_{mn}}{\partial x_r} dx_m\, dx_n\, \delta x_r + 2 g_{mn}\, dx_m\, d\delta x_n \right\} = 0,$$

d and δ having been interchanged in the last term in the bracket. We thus obtain

$$\int_1^2 \frac{1}{2\,ds} \left\{ \frac{\partial g_{mn}}{\partial x_r} dx_m\, dx_n\, \delta x_r + d(2 g_{mn}\, dx_m\, \delta x_n) - \delta x_n d(2 g_{mn}\, dx_m) \right\} = 0.$$

The middle term contributes nothing to the integral, just as in

the case of the corresponding term in **(35·721)**, since the δx_n vanish at both limits, and we get

$$\int_1^2 ds\ \delta x_r \left\{ \tfrac{1}{2}\frac{\partial g_{mn}}{\partial x_r}\frac{dx_m}{ds}\frac{dx_n}{ds} - \frac{\partial g_{mr}}{\partial x_n}\frac{dx_m}{ds}\frac{dx_n}{ds} - g_{pr}\frac{d^2x_p}{ds^2}\right\} = 0,$$

in which certain permissible and easily intelligible changes have been made in the suffixes.

Since δx_r is arbitrary, we infer that

$$g_{rp}\frac{d^2x_p}{ds^2} + \left(\frac{\partial g_{mr}}{\partial x_n} - \tfrac{1}{2}\frac{\partial g_{mn}}{\partial x_r}\right)\frac{dx_m}{ds}\frac{dx_n}{ds} = 0.$$

Interchanging m and n and remembering that g_{mn} is symmetrical, we get

$$g_{rp}\frac{d^2x_p}{ds^2} + \left(\frac{\partial g_{nr}}{\partial x_m} - \tfrac{1}{2}\frac{\partial g_{mn}}{\partial x_r}\right)\frac{dx_m}{ds}\frac{dx_n}{ds} = 0;$$

and, on adding and dividing by 2, we obtain finally:

$$g_{rp}\frac{d^2x_p}{ds^2} + \tfrac{1}{2}\left(\frac{\partial g_{mr}}{\partial x_n} + \frac{\partial g_{nr}}{\partial x_m} - \frac{\partial g_{mn}}{\partial x_r}\right)\frac{dx_m}{ds}\frac{dx_n}{ds} = 0. \quad \textbf{(35·77)}$$

These equations, which describe a rectilinear path in a Galilean continuum and of which there are four ($r = 1, 2, 3, 4$), may be put in a quite different and, as we shall see later, very suggestive form by giving a stationary value to **(35·71)**—which obviously amounts to the same thing as **(35·72)**—and making use of Stokes' theorem. That is to say we start out from

$$\delta\int_1^2 p_\alpha dx_\alpha = 0,$$

which may be written

$$\oint p_\alpha dx_\alpha = 0, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(35·78)}$$

where the integral is taken round the closed loop made up of the varied path from 1 to 2 and the actual path in the reverse direction, i.e. from 2 to 1. On applying Stokes' theorem to this equation we get

$$\tfrac{1}{2}\iint\left\{\frac{\partial p_\alpha}{\partial x_\beta} - \frac{\partial p_\beta}{\partial x_\alpha}\right\}(\delta x_\beta\, dx_\alpha - \delta x_\alpha\, dx_\beta) = 0.$$

Each surface element, $ds^{\beta\alpha} \equiv \delta x_\beta\, dx_\alpha - \delta x_\alpha\, dx_\beta$, is of course arbitrary, since every δx is arbitrary. Hence

$$\frac{\partial p_\alpha}{\partial x_\beta} - \frac{\partial p_\beta}{\partial x_\alpha} = 0,$$

and consequently also

$$\left(\frac{\partial p_\alpha}{\partial x_\beta} - \frac{\partial p_\beta}{\partial x_\alpha}\right)\frac{dx_\beta}{d\tau} = 0. \quad . \quad . \quad . \quad \textbf{(35·781)}$$

There are four of these equations ($\alpha = 1, 2, 3, 4$) and they are simply equations **(35·77)** in another form.

So far we have been studying a path which is 'straight' and the last equation (or set of equations) reduces, when we use orthogonal (rectangular) co-ordinates, to

$$\frac{\partial p_\alpha}{\partial x_\beta}\frac{dx_\beta}{d\tau} = \frac{dp_\alpha}{d\tau} = 0,$$

the contribution

$$-\frac{\partial p_\beta}{\partial x_\alpha}\frac{dx_\beta}{d\tau}$$

vanishing with such co-ordinates.

§ 35·8. A Digression on the Dynamical Equations of Hamilton and Lagrange

The discussion in the preceding section makes it obvious that the dynamical equations of Hamilton and Lagrange continue in force for a particle within the wider scope of the special theory of relativity. It is instructive to verify this. The Hamiltonian function, H, is

$$H = m_0c^2\gamma + V \quad . \quad . \quad . \quad . \quad . \quad \textbf{(35·8)}$$

provided γ is expressed as a function of the components of the momentum of the particle. V is a function of the co-ordinates and the time with the property that the Newtonian force exerted on the particle is equal to $-$ **grad** V.

Now

$$\frac{1}{\gamma^2} = 1 - \frac{\mathbf{v}^2}{c^2}.$$

Hence

$$\mathbf{p}^2 = p_x^2 + p_y^2 + p_z^2 = m_0^2c^2(\gamma^2 - 1).$$

Consequently

$$\gamma = \left\{\frac{p_x^2 + p_y^2 + p_z^2}{m_0^2c^2} + 1\right\}^{1/2}$$

This has to be substituted for γ in **(35·8)**. Thus we get

$$H = \{(p_x^2 + p_y^2 + p_z^2)c^2 + m_0^2c^4\}^{1/2} + V \quad \textbf{(35·81)}$$

and therefore

$$\frac{dx}{dt} = \frac{\partial H}{\partial p_x},$$

and there are corresponding expressions for the remaining velocity components. The definition of V of course implies that

$$\frac{dp_x}{dt} = -\frac{\partial V}{\partial x},$$

and therefore

$$\frac{dp_x}{dt} = -\frac{\partial H}{\partial x},$$

and there follow similar equations for the remaining components.

The generalization of Hamilton's principal function is

$$\int (p_x\,dx + p_y\,dy + p_z\,dz - E\,dt),$$

which we may write

$$\int L\,dt$$

so that we get for the Lagrangian function, L,

$$L = p_x\frac{dx}{dt} + p_y\frac{dy}{dt} + p_z\frac{dz}{dt} - E,$$

or

$$L = mv^2 - mc^2 - V$$

and therefore

$$L = -m_0c^2\gamma^{-1} - V,$$

or finally

$$L = -m_0c^2\{1 - (v_x^2 + v_y^2 + v_z^2)/c^2\}^{1/2} - V. \quad . \quad \textbf{(35·82)}$$

Hence

$$\frac{\partial L}{\partial v_x} = m_0\gamma v_x = p_x.$$

Now

$$\frac{dp_x}{dt} = -\frac{\partial V}{\partial x},$$

and therefore

$$\frac{d}{dt}\left(\frac{\partial L}{\partial v_x}\right) = \frac{\partial L}{\partial x},$$

and we have, of course, two further equations similarly related to the Y and Z axes.

We have been assuming rectangular axes of co-ordinates so far; but it is easy to verify that the validity of the equations of Hamilton and Lagrange is not confined to them. We shall content ourselves with a verification of this for the case of polar co-ordinates. Equation **(35·81)** will now take the form

$$H = (p^2c^2 + m_0^2c^4)^{1/2} + V,$$

where

$$p^2 = p_r^2 + p_\theta^2/r^2 + p_\phi^2/r^2\sin^2\theta.$$

Hence, for example,

$$p\frac{\partial p}{\partial p_\theta} = \frac{p_\theta}{r^2}$$

and therefore

$$\frac{\partial H}{\partial p_\theta} = c^2 p \frac{\partial p}{\partial p_\theta} / (p^2c^2 + m_0{}^2c^4)^{1/2} = \frac{p_\theta}{mr^2},$$

or

$$\frac{\partial H}{\partial p_\theta} = \frac{d\theta}{dt}.$$

All this follows really from Hamilton's principle, which, as we have learned in the last section, continues in force under special relativity provided we define momentum **(p)**, and energy in the way we have done and the present section merely contains verifications of this.

§ 35·9. The General Principle of Relativity

The interval form **(35·76)** suggests possibilities of further progress. In the first place it is associated with greater freedom in the choice of co-ordinate systems, so that we may essay a correspondingly more general form of the **principle of relativity** which, in effect, in the special theory **(§ 34·5)** has been confined to spacial co-ordinate systems moving relatively to one another with constant velocities. We shall now lay down as a fundamental principle that the equations describing physical phenomena have the same form in all co-ordinate systems with the corollary that they must be expressions of tensor relationships. This is Einstein's **general principle of relativity.** In the second place we may contemplate a still more general type of 4-dimensional continuum—a **Riemannian continuum** [1]—which includes the Galilean one as a special case. It is characterized by the property that the square of the interval between two neighbouring points has the form **(35·76).** But it is not in general possible to find co-ordinate systems which enable the square of the interval to take the form

$$ds^2 = \sum_1^4 dx_\alpha{}^2 \quad . \quad . \quad . \quad . \quad . \quad (35·9)$$

everywhere in the continuum. The Galilean continuum is the special case of the Riemannian one which emerges when such co-ordinate systems exist which give the square of the interval the form **(35·9)** everywhere in the continuum.

[1] So named after G. F. B. Riemann (1826–66), a great German mathematician, famous for his contributions to the theory of analytic functions and for the special form he gave the theory. Riemann surfaces have no relationship to our geometrical studies. They were part of the symbolism of his theory of functions.

§ 36. The Riemannian Continuum and the Principle of Equivalence

In the Riemannian continuum equations **(35·77)** and **(35·781)** represent something which is a generalization of a straight line, namely a **geodesic**—a shortest or a longest line, for example. The geodesic is in fact a straight line in one special case, namely, when a system of co-ordinates can be found for which **(35·9)** is true. The continuum in this case is Galilean. The Galilean continuum is however only a special case of the more general Riemannian one. The surface of a sphere furnishes us with a 2-dimensional illustration of a Riemannian continuum. In this case there is no system of co-ordinates for which **(35·9)** holds. The geodesics in this 2-dimensional continuum are great circles. A sufficiently small part of the spherical surface can be regarded, practically, as part of a plane surface (a tangential plane), and we can find co-ordinate systems which, *within such a restricted region*, give the square of the interval the form **(35·9)**. For example, we might choose as the co-ordinates of a point the distance (x) measured along an equator from some chosen point on it (origin) to the circle of longitude on which the point in question lies and the distance (y) measured along this circle from the equator to the point. If C represents the circumference of the sphere, the possible values of the first co-ordinate (x) lie between $+C/2$ and $-C/2$, while those of the second co-ordinate (y) lie between $+C/4$ and $-C/4$. In the neighbourhood of the equator, and only there, the expression for the square of the interval approaches the limit

$$ds^2 = dx^2 + dy^2.$$

So too in a 4- or more-dimensional Riemannian continuum, though we cannot in general find co-ordinate systems which give ds^2 the form **(35·9)** *everywhere*, we can choose them in such a way that **(35·9)** will be true in some small neighbourhood which we may regard as common to the (curved) Riemannian continuum and a tangential (plane) Galilean one.

The question now arises: What physical significance can be assigned to the geodesics? An answer to this question is suggested when we remember that the geodesics are practically straight lines in a sufficiently restricted neighbourhood and when we remember further one of the most striking and well-established of all the facts about gravitation, namely, that all free bodies in a small neighbourhood move with the same gravitational acceleration when their mutual influence on one another can be neglected. Near the earth's surface, for example, it is the familiar g of the elementary textbooks and is independent of the mass of the body

and of the nature of the material constituting it. Now the generalized principle of relativity **(§ 35·9)** authorizes us to employ any system of reference and assures us that the equations describing physical phenomena have the same form, whatever system of reference may be used. Consider, then, a body moving under the influence of a gravitational field only, for example, that near the earth's surface, and let its motion be referred to a freely falling structure in the same neighbourhood, as a system of reference; or let it be observed by a freely falling observer. The important fact about gravitation just described makes it manifest that the body in question will pursue **a rectilinear path, as seen by such an observer, so long as the observer and it are in the same sufficiently small neighbourhood.** Our co-ordinate system, fixed in the freely falling structure, can be so chosen that the square of the interval is expressed by **(35·9)**. The equations of motion of the particle, using such co-ordinates, will therefore be

$$\frac{d^2x_\alpha}{ds^2} = 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (36)$$

and they are equivalent, as we know, to

$$\delta\int_1^2 ds = 0, \quad . \quad . \quad . \quad . \quad . \quad (36{\cdot}01)$$

which is quite independent of the choice of a system of reference. **This is true so long as we confine our attention to a sufficiently restricted neighbourhood.** It is simply an expression of observational fact. Outside this neighbourhood, i.e. when the body or particle gets far away from the observer, or from the freely falling structure in which the co-ordinate origin and axes are fixed, its motion will cease to be rectilinear. In other words, equations **(36)** will no longer describe it. It is conceivable, however, that **(36·01)** which, as an observational fact, correctly describes the motion of the particle in a restricted neighbourhood, may correctly describe it always; since in a Riemannian continuum geodesics need not be straight lines, though undistinguishable from them in a sufficiently small part of it. If this be true, the equations of motion of a particle in a gravitational field are the equations **(35·77)** and its path is a geodesic. We shall adopt this supposition and raise it to the rank of a fundamental principle. It is known as the **principle of equivalence** and so called because it regards motion in a gravitational field, when this deviates from rectilinearity, as of the same nature as deviations from rectilinearity which are due to changing the system of reference. It

is a kind of extrapolation and its adoption will be justified by the success attending it.

It should be noted that the Minkowskian force **(§ 35)** vanishes in a gravitational field. This must be the case since its r component is equal to the rest mass multiplied by the expression on the left of **(35·77)**.

§ 36·1. Electromagnetic Phenomena

In Einstein's gravitational theory, as we have just seen, the motion of a particle (in a gravitational field) is the exact parallel or analogue of the Newtonian uniform motion of a particle along a straight line. The expression on the left-hand side of **(35·77)**, or of **(35·781)**, is a generalization of the Newtonian d^2x/dt^2, etc., or of dp_x/dt, etc.; or, as we might still better express it, a generalization of the Minkowskian force $m_0 d^2x/d\tau^2$, etc. There is indeed no necessity for the mass of the particle to appear at all (indeed, it does not appear in **35·77**) since the extended force which appears in these equations vanishes. Of course there is one way in which masses will make their appearance: if there were no masses there would be no permanent gravitational fields, and the geodesics would become straight lines. Evidently the shapes of the geodesics are determined by masses and the law of gravitation, which we shall have to approach later, must therefore be a law which determines for us the coefficients, g_{mn}. It must be, in fact, a law which dictates some kind of limitation to the Riemannian space-time of Einstein's theory. Briefly, masses, i.e. proper masses, are going to become constants which appear in the description of the structure of space-time and **gravitational phenomena are the expression of the geometrical features of the continuum.** They are an inherent part of the continuum and are not to be regarded as something distinct from it as an actor is from the stage on which he plays his part. Electromagnetic phenomena, however, and indeed all phenomena which are not gravitational, are not included in the description of space-time in Einstein's theory. For such phenomena there is indeed a place in the 4-dimensional Riemannian continuum; but they are merely housed by it, as it were, and are not an organic part of it.

A question now arises naturally. Is it not possible in some way to widen our theory so that electromagnetic, just as much as gravitational phenomena, may become an expression of the geometry, or of the metrical features of the continuum? There is little doubt that this is the case. A number of very different suggestions have already been made to achieve this kind of

unification. The earliest of them was described by Weyl and elaborated by Eddington. It is doubtful whether it is the right suggestion for dealing with electromagnetic phenomena; but it is of great interest and the fundamental notion contained in it may prove of service in another direction. Einstein himself described a unitary theory of great interest; but this too does not appear to furnish an acceptable solution of the problem of making electromagnetic phenomena an organic outcome of the geometrical properties of the continuum. The most attractive and probably the correct solution is one which has been developed by Kaluza and others. It is possible to indicate its nature here. It is a natural extension of Einstein's general theory and can be developed by starting out from the geodesic equations **(35·781)** combined with what we learned about the motion of a charged particle in an electromagnetic field while studying Einstein's special theory.

Using orthogonal co-ordinates, i.e. co-ordinates so chosen that the square of the interval is expressed by **(35·9)**,[1] we have for the X component of the (Minkowskian) force on a charged particle in an electromagnetic field:

$$\mathscr{F}_x = 0 \times \frac{4\pi Ae}{a}\frac{dx}{d\tau} + \frac{\mu H_z}{4\pi A} \times \frac{4\pi Ae}{a}\frac{dy}{dz} + \left(\frac{-\mu H_y}{4\pi A}\right) \times \frac{4\pi Ae}{a}\frac{dz}{d\tau} + \left(\frac{-i\sqrt{\mu K}}{4\pi A}\mathcal{E}_x\right)\frac{4\pi Ae}{a}\frac{dw}{dt}.$$

or $$\mathscr{F}_1 = \frac{4\pi Ae}{a}\left\{F_{11}\frac{dx_1}{d\tau} + F_{12}\frac{dx_2}{d\tau} + F_{13}\frac{dx_3}{d\tau} + F_{14}\frac{dx_4}{d\tau}\right\},$$

which may be written:

$$\mathscr{F}_\alpha = \frac{4\pi Ae}{a}\left\{\frac{\partial A_\beta}{\partial x_\alpha} - \frac{\partial A_\alpha}{\partial x_\beta}\right\}\frac{dx_\beta}{d\tau}.$$

We might of course adopt units for which $4\pi A/a = 1$. Let us instead use the letter **K** in the sense

$$K_\alpha = (4\pi A/a)A_\alpha. \quad . \quad . \quad . \quad . \quad \textbf{(36·1)}$$

Then we have

$$\mathscr{F}_\alpha = \left\{\frac{\partial(eK_\beta)}{\partial x_\alpha} - \frac{\partial(eK_\alpha)}{\partial x_\beta}\right\}\frac{dx_\beta}{d\tau}. \quad . \quad . \quad \textbf{(36·101)}$$

Where $e\mathbf{K}$ is a vector and the bracket expression **(curl of eK)** is a tensor whatever sort of co-ordinates we may use, as we shall learn later. Now a general expression for $\mathscr{F}_\alpha$, the Minkowskian

[1] We may do this in a small neighbourhood.

force on a particle, which holds in all co-ordinate systems, is **(35·781)**,

$$\mathscr{F}_\alpha = \left\{\frac{\partial p_\alpha}{\partial x_\beta} - \frac{\partial p_\beta}{\partial x_\alpha}\right\}\frac{dx_\beta}{d\tau}.$$

Consequently **(36·101)** may be written:

$$\left\{\frac{\partial}{\partial x_\beta}(p_\alpha + eK_\alpha) - \frac{\partial}{\partial x_\alpha}(p_\beta + eK_\beta)\right\}\frac{dx_\beta}{d\tau} = 0,$$

or

$$\left\{\frac{\partial \Pi_\alpha}{\partial x_\beta} - \frac{\partial \Pi_\beta}{\partial x_\alpha}\right\}\frac{dx_\beta}{d\tau} = 0, \quad . \quad . \quad \textbf{(36·11)}$$

in which **Π** is a new vector defined by

$$\Pi_\alpha = p_\alpha + eK_\alpha.^1 \quad . \quad . \quad . \quad . \quad \textbf{(36·12)}$$

Indeed, we may represent the motion of a charged particle in an electromagnetic field by an extended form of Hamilton's (or Maupertius') principle, namely,

$$\delta\int_1^2 \Pi_\alpha dq_\alpha = 0.$$

(See **35·71** and compare **35·781** with **36·11.**)

It will be observed that the equations **(36·11)** have precisely the same form as those of a geodesic (cf. **35·781**); but they are evidently not those of a geodesic in the 4-dimensional Riemannian continuum of Einstein's general theory, except of course in the very special case where the vector potential, **K,** vanishes or is constant. Can we enlarge our continuum, or in some way widen our theory, so that the equations **(36·11)** will actually be those of a geodesic? In the Galilean space-time continuum $p_\alpha dx_\alpha$ is an invariant. In terms of orthogonal co-ordinates it is

$$p_\alpha dx_\alpha = m_0\frac{\Sigma dx_\alpha{}^2}{d\tau},$$

and its general expression is therefore

$$p_\alpha dx_\alpha = m_0 ds^2/d\tau = m_0 g_{\alpha\beta}\, dx_\alpha\, dx_\beta/d\tau,$$

so that the general expression for p_α is

$$p_\alpha = m_0 g_{\alpha\beta}\, dx_\beta/d\tau \quad \text{(cf. § 8·2)}.$$

Consequently we obtain for the new vector **Π** the expression:

$$\Pi_\alpha = m_0 g_{\alpha\beta}\, dx_\beta/d\tau + eK_\alpha. \quad . \quad . \quad \textbf{(36·121)}$$

The momentum **Π** is therefore a vector each of whose components is equal to the corresponding one of **p** added to the product of the same component of the vector potential and the

[1] W. Wilson: *Proc. Roy. Soc.*, *A*, 102, p. 478 (1922).

charge on the particle. Noting this and also the identity in form of **(35·781)** and **(36·11),** we are naturally tempted to regard **Π** as a (covariant) vector in a 5-dimensional continuum of a Riemannian type. This at once raises the question as to the significance of the fifth component, Π_5, and it would be rather unsatisfactory and disappointing if we were unable to find an observational equivalent for it. We have not the same anxiety about the fifth co-ordinate, x_5. If no observational equivalent turns up for that we may place it in the category of cyclic co-ordinates. Happily there is a most important physical quantity which can be identified with Π_5, and the identification is the most natural and appropriate one imaginable. Among the fundamental principles which have guided us in the development of physical theory, including relativity, up to this point is that of conservation of momentum; which asserts that Σp_α remains constant for a number of particles subject only to their mutual interactions. Naturally we must widen its scope so that $\Sigma\Pi_\alpha$ is a constant. For $\alpha = 1, 2, 3, 4$ this will cause us no difficulties and will include the conservation of momentum in the older and more restricted sense as well as that of mass and energy. Is there any further physical quantity which is conserved and which has not already been subsumed under the broad principle of conservation of momentum? There is the **conservation of electric charge.** Therefore if e represent the charge on a particle, let us write

$$e = a_0\Pi_5, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(36·13)}$$

where a_0 is a suitable constant. This not only disposes of Π_5, but it unifies all the conservation principles in the one great principle of conservation of momentum.

The motion of a particle, charged or not charged and whether in a gravitational or electromagnetic field or in both, will be described by the five equations **(36·11)** or by five equations exactly like **(35·77)**, except that the tensor, g_{mn}, must be replaced by a corresponding tensor, $\gamma_{\alpha\beta}$, appropriate to the 5-dimensional continuum. Either of these sets of equations describes a geodesic in the continuum and is equivalent to

$$\delta\int_1^2 d\sigma = 0, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(36·14)}$$

$d\sigma$ being the interval, i.e.

$$d\sigma^2 = \gamma_{\alpha\beta}\, dx_\alpha\, dx_\beta. \quad . \quad . \quad . \quad . \quad \textbf{(36·15)}$$

Let us now imagine the simplest case of a charged particle. This will be a particle moving in a region where there is no field

of force of any kind. Our new 5-dimensional continuum will therefore acquire a Euclidean or Galilean character, i.e. it will be possible to find co-ordinate systems in which the square of the interval, $d\sigma$, takes the form:

$$d\sigma^2 = dx_1^2 + dx_2^2 + dx_3^2 + dx_4^2 + dx_5^2 = ds^2 + dx_5^2, \quad \textbf{(36·151)}$$

and, of course, when $\alpha = 1, 2, 3$ or 4, Π_α will become p_α. Therefore

$$\Pi_\alpha = m\frac{dx_\alpha}{dt} = m_0\frac{dx_\alpha}{d\tau} = \mu\frac{dx_\alpha}{dT}, \quad \textbf{(36·152)}$$

the last term anticipating a new proper time, T, and proper mass, μ, appropriate to the new continuum. Then further

$$\Pi_5 = \mu\frac{dx_5}{dT}. \quad \textbf{(36·153)}$$

Remembering that the old proper time was defined (cf. **§ 34·9**) by

$$ds^2 = -c^2d\tau^2,$$

we shall define the new one by

$$d\sigma^2 = -c^2dT^2, \quad \textbf{(36·16)}$$

so that $$-c^2dT^2 = -c^2d\tau^2 + dx_5^2$$

and $$\frac{dT}{d\tau} = \left\{1 - \frac{1}{c^2}\left(\frac{dx_5}{d\tau}\right)^2\right\}^{\frac{1}{2}},$$

or

$$d\tau = \gamma' dT, \quad \textbf{(36·161)}$$

where

$$\gamma' = (1 - v'^2/c^2)^{-1/2} \quad \textbf{(36·162)}$$

and

$$v' = dx_5/d\tau. \quad \textbf{(36·163)}$$

The new proper mass, μ, will be related to the old one, m_0, as will be easily understood on referring again to **§ 35**, by the equation

$$m_0 = \gamma'\mu, \quad \textbf{(36·17)}$$

and of course we have further

$$m/dt = m_0/d\tau = \mu/dT. \quad \textbf{(36·171)}$$

We have anticipated **(34·6)** that we shall have to regard light (the phenomenon usually described as electromagnetic waves) rather as Newton did; that is to say, we shall find that we must regard it as made up of something resembling particles —they are termed **photons**—and possessing the property or peculiarity that Newton described as **fits**. This does not imply that the electromagnetic theory we have described at such length

is going to be discarded. It means in the main that the interpretation of it will be changed.

Photons travel of course with the velocity c. Therefore ds vanishes along the track of a photon, and so does $d\tau$ in consequence. The α component of momentum of a photon is for $\alpha = 1, 2, 3, 4$

$$m\,dx_\alpha/dt = m_0\,dx_\alpha/d\tau = \mu\,dx_\alpha/dT,$$

consequently, though m is not vanishingly small, $m_0 = 0$.

Further, since photons are uncharged

$$\Pi_5 = m\,dx_5/dt = m_0\,dx_5/d\tau = \mu\,dx_5/dT = 0,$$

and as $m/dt = m_0/d\tau = \mu/dT$ does not vanish, we must conclude that dx_5 vanishes for a photon and indeed for any uncharged particle. Therefore not only does ds vanish for a photon; but also $d\sigma$. And parallel with this m_0 and μ vanish for a photon.

Imagine instead of a photon any charged particle for which of course m_0 and ds do not vanish. It is a permissible and interesting speculation to contemplate the hypothesis that nevertheless μ and $d\sigma$ vanish for all elementary particles. If $d\sigma = 0$, then

$$ds^2 + dx_5^2 = 0,$$

and consequently

$$v'^2 = (dx_5/d\tau)^2 = -(ds/d\tau)^2 = c^2.$$

Hence

$$e = a_0\Pi_5 = a_0 m_0\,dx_5/d\tau = \pm\, a_0 m_0 c.$$

We naturally assign a positive value to a_0 and therefore

$$a_0 = |\,e/m_0 c\,|. \quad . \quad . \quad . \quad . \quad \textbf{(36·18)}$$

If the very speculative hypothesis we are now studying, namely, that $d\sigma = 0$ for all elementary particles, electrons, positrons, photons, be a correct one we must conclude that e/m_0 has the same value for all of them and all of them have the proper mass $\mu = 0$. The photon is merely the special case where m_0 (and consequently e) vanish. We shall leave this part of the subject of relativity, i.e. the problem of the unification of gravitational and electromagnetic phenomena, at this point and continue the study of relativity as it appears in its Minkowski-Einstein guise.

CHAPTER V

TENSOR ANALYSIS

§ 36·2. The Fundamental Tensor

WE shall now develop the tensor analysis appropriate for a continuum, Galilean or other, in which the square of the interval has the form:

$$ds^2 = g_{mn}\, dx_m\, dx_n. \quad . \quad . \quad . \quad . \quad \textbf{(36·2)}$$

For a 3-dimensional continuum for example

$$\begin{aligned} ds^2 = {} & g_{11}\, dx_1{}^2 + g_{12}\, dx_1\, dx_2 + g_{13}\, dx_1\, dx_3 \\ & + g_{21}\, dx_2\, dx_1 + g_{22}\, dx_2\, dx_2 + g_{23}\, dx_2\, dx_3 \\ & + g_{31}\, dx_3\, dx_1 + g_{32}\, dx_3\, dx_2 + g_{33}\, dx_3{}^2. \end{aligned} \quad . \quad . \quad \textbf{(36·201)}$$

We shall always take g_{mn} to be symmetrical, i.e. $g_{mn} = g_{nm}$. We are at liberty to do this, since it is equivalent to defining either g_{mn} or g_{nm} as one half of the coefficient of the product $dx_m\, dx_n$ in the sum which makes up $\mathbf{ds}^2$. The term containing $dx_m\, dx_n$ is split into two equal parts, $g_{mn}\, dx_m\, dx_n$ and $g_{nm}\, dx_n\, dx_m$—the summation convention is not now implied—for the sake of the great advantage of being able to represent the whole sum $\mathbf{ds}^2$ in the simple form $g_{mn}\, dx_m\, dx_n$ when the summation convention is used.

The expression **(36·2)** is, by hypothesis, an invariant of our transformations, and we can prove that the g_{mn} constitute a covariant tensor.

It follows from **(36·2)** that

$$g_{mn}(dx_m + \delta x_m)(dx_n + \delta x_n)$$

is an invariant, $\mathbf{dx}$ and $\boldsymbol{\delta}\mathbf{x}$ being two arbitrary (displacement) vectors.

Hence

$$g_{mn}\, dx_m\, \delta x_n$$

is an invariant, and since $dx_m\, \delta x_n$ is the mn component of an arbitrary contravariant tensor of rank two, it follows that g_{mn} is a covariant tensor of rank two **(§§ 34·7** and **34·8).**

Let us consider next the determinant:

$$g \equiv |\, g_{mn}\,|. \quad . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(36·21)}$$

It may be written:

$$g = \sum_n g_{mn}\Delta^{mn},$$

in which *the summation is extended over the index n only.* Δ^{mn} is a minor determinant obtained by omitting from $|\, g_{mn}\,|$ the row, m, and the column, n, and giving the result a suitable sign. We may illustrate this by one of the expressions for g in a 3-dimensional continuum:

$$g = \begin{vmatrix} g_{11}, & g_{12}, & g_{13} \\ g_{21}, & g_{22}, & g_{23} \\ g_{31}, & g_{32}, & g_{33} \end{vmatrix}$$

One of its expansions, for example, is

$$g_{31}\begin{vmatrix} g_{12}, & g_{13} \\ g_{22}, & g_{23} \end{vmatrix} + g_{32}\left\{ - \begin{vmatrix} g_{11}, & g_{13} \\ g_{21}, & g_{23} \end{vmatrix}\right\} + g_{33}\begin{vmatrix} g_{11}, & g_{12} \\ g_{21}, & g_{22} \end{vmatrix}.$$

Thus, for example,

$$\Delta^{32} = - \begin{vmatrix} g_{11}, & g_{13} \\ g_{21}, & g_{23} \end{vmatrix}.$$

It is obvious that

$$g_{an}\Delta^{bn} = 0, \text{ when } a \neq b,$$

because it is identical with what is obtained by replacing in the determinant, g, the row, b, by a row identical with a. Therefore

$$g_{an}\Delta^{bn} = \begin{cases} g, & a = b, \\ o, & a \neq b, \end{cases}$$

the summation in either case being with respect to n. Hence

$$g_{an}\{\Delta^{bn}/g\} = \begin{cases} 1, & a = b, \\ 0, & a \neq b. \end{cases}$$

Or, if we define g^{bn} by

$$g^{bn} \equiv \Delta^{bn}/g, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(36·22)}$$

we have

$$g_a{}^b \equiv g_{an}g^{bn} = \begin{cases} 1, & a = b, \\ 0, & a \neq b. \end{cases} \quad . \quad . \quad . \quad \textbf{(36·23)}$$

The set of quantities $g_a{}^b$ constitutes a mixed tensor of the second rank; because

$$g_a{}^b B^a = B^b,$$

where B^a is an arbitrary tensor (**§§ 34·7** and **34·8**). We may also show that the set of quantities g^{am} constitutes a contravariant tensor of rank two. We have, namely,

$$g^{am}g_{bm}B_a{}^b \equiv \textit{Invariant}$$

since it is equal to

$$g_b{}^a B_a{}^b,$$

where $B_a{}^b$ is *any* tensor of the kind described by the notation. Therefore

$$g^{am}B_{ma} \equiv \textit{Invariant}$$

where B_{ma} is the tensor $g_{bm}B_a{}^b$. Now this latter tensor is arbitrary

on account of the arbitrariness of B_a^b. Hence g^{am} is a tensor. The term **fundamental tensor** is applied more particularly to the tensor g_{mn}; but g^{mn} and other tensors, including the very important Riemann-Christoffel tensor of **§ 36·8,** which can be expressed in terms of g_{mn} and its derivatives, are often called fundamental tensors.

We may illustrate some features of the more extended tensor calculus we are now studying by the example of polar co-ordinates in Euclidean space. The typical contravariant vector is now

$$(dr,\ d\theta,\ d\phi),$$

or

$$(dx_1,\ dx_2,\ dx_3),$$

and since

$$ds^2 = dr^2 + r^2\,d\theta^2 + r^2\sin^2\theta\,d\phi^2,$$

or, as we may otherwise express it,

$$ds^2 = dx_1^2 + x_1^2\,dx_2^2 + x_1^2\sin^2 x_2\,dx_3^2,$$

therefore

$$g_{11} = 1,\ g_{22} = x_1^2,\ g_{33} = x_1^2\sin^2 x_2,$$

while all the other g_{mn} vanish.

Now $g_{mn}\,dx_n$ is a covariant vector; its three components in the present illustration being:

$$g_{11}\,dx_1 + g_{12}\,dx_2 + g_{13}\,dx_3,$$
$$g_{21}\,dx_1 + g_{22}\,dx_2 + g_{23}\,dx_3,$$

and

$$g_{31}\,dx_1 + g_{32}\,dx_2 + g_{33}\,dx_3,$$

or

$$dr,\ r^2\,d\theta \text{ and } r^2\sin^2\theta\,d\phi.$$

In Euclidean space and Newtonian mechanics, mass and time are invariants. Therefore we have as an example of a covariant vector

$$\left(m\frac{dr}{dt},\ mr^2\frac{d\theta}{dt},\ mr^2\sin^2\theta\frac{d\phi}{dt}\right),$$

which is the familiar generalized momentum of a particle in polar co-ordinates.

§ 36·3. The Geodesic Equation

In equation **(35·77)** let us write

$$\Gamma_{mn,\,r} \equiv \frac{1}{2}\left(\frac{\partial g_{mr}}{\partial x_n} + \frac{\partial g_{nr}}{\partial x_m} - \frac{\partial g_{mn}}{\partial x_r}\right), \quad . \quad . \quad \textbf{(36·3)}$$

so that the equation becomes

$$g_{pr}\frac{d^2x_p}{ds^2} + \Gamma_{mn,\,r}\frac{dx_m}{ds}\frac{dx_n}{ds} = 0.$$

Multiply both sides by g^{rs}, summing with respect to r, thus

$$g^{rs}g_{pr}\frac{d^2x_p}{ds^2} + g^{rs}\Gamma_{mn,\,r}\frac{dx_m}{ds}\frac{dx_n}{ds} = 0.$$

Therefore equations **(35·77)** take the form

$$\frac{d^2x_s}{ds^2} + \Gamma^s_{mn}\frac{dx_m}{ds}\frac{dx_n}{ds} = 0, \quad . \quad . \quad . \quad \textbf{(36·301)}$$

where

$$\Gamma^s_{mn} \equiv g^{rs}\Gamma_{mn,\,r}. \quad . \quad . \quad . \quad . \quad \textbf{(36·302)}$$

The equations **(36·301)** constitute therefore an alternative form of the equations of a geodesic which, as we have seen, is the sort of path along which a particle in a gravitational field (and not subject to other forces) must travel; and the problem that remains, so far as gravitation is concerned, is that of determining the components of the tensor g_{mn}. This we shall leave to a later section.

§ 36·4. The Determinant g

The components $g_{\mu\nu}'$ of the fundamental tensor in one system of co-ordinates are related, as we know, to its components g_{mn} in another by the equations

$$g_{\mu\nu}' = \frac{dx_m}{dx_\mu'}\frac{dx_n}{dx_\nu'}g_{mn},$$

so that the determinant

$$g' \equiv |\, g_{\mu\nu}' \,|$$

may be expressed by

$$g' = \left| \frac{\partial x_m}{dx_\mu'}\frac{dx_n}{dx_\nu'}g_{mn} \right|,$$

the typical constituent, $g_{\mu\nu}'$, having been replaced by the equivalent expression

$$\frac{\partial x_m}{dx_\mu'}\frac{dx_n}{dx_\nu'}g_{mn},$$

which is a sum of products, each containing three factors. These products include for example

$$\frac{\partial x_1}{\partial x_\mu'}\frac{\partial x_1}{\partial x_\nu'}g_{11}, \qquad \frac{\partial x_1}{\partial x_\mu'}\frac{\partial x_2}{\partial x_\nu'}g_{12},$$

and so on. It is helpful here to introduce, temporarily, a slightly different notation. We shall represent

$$\partial x_n/\partial x_\nu' \text{ by } k_{n\nu},$$
$$\partial x_m/\partial x_\mu' \text{ by } l_{\mu m}.$$

The typical constituent of the determinant thus becomes

$$l_{\mu m} g_{mn} k_{n\nu},$$

summation with respect to m and n being implied. Finally let us represent

$$g_{mn} k_{n\nu} \text{ by } K_{m\nu},$$

the summation with respect to n having been effected. Thus

$$g' = | \, l_{\mu m} K_{m\nu} \, |.$$

Consequently

$$g' = | \, l_{\mu m} \, | \times | \, K_{m\nu} \, |$$

by the rule for multiplying determinants and

$$| \, K_{m\nu} \, | = | \, g_{mn} \, | \times | \, k_{n\nu} \, |.$$

Hence

$$g' = | \, l_{\mu m} \, | \times | \, k_{n\nu} \, | \times | \, g_{mn} \, |$$

and consequently

$$g' = \left| \frac{\partial x_m}{\partial x_\mu'} \right| \times \left| \frac{\partial x_n}{\partial x_\nu'} \right| \times \left| g_{mn} \right|;$$

so that finally

$$\left.\begin{aligned} g' &= J^2 g, \\ \text{and} \qquad g'^{1/2} &= J g^{1/2}, \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad \textbf{(36·4)}$$

where

$$J \equiv \left| \frac{\partial x_m}{\partial x_\mu'} \right| = \left| \frac{\partial x_n}{\partial x_\nu'} \right| \equiv \frac{\partial (x_1 \ . \ . \ . \ x_4)}{\partial (x_1' \ . \ . \ . \ x_4')}$$

is the familiar Jacobean determinant. Now we know that

$$J \, dx_1' \, dx_2' \, dx_3' \, dx_4' = dx_1 \, dx_2 \, dx_3 \, dx_4,^1$$

and on multiplying this by **(36·4)** we get

$$g'^{1/2} \, dx_1' \, dx_2' \, dx_3' \, dx_4' = g^{1/2} \, dx_1 \, dx_2 \, dx_3 \, dx_4 \quad . \quad . \quad . \quad \textbf{(36·41)}$$

so that the product of $g^{1/2}$ and $dx_1 \, dx_2 \, dx_3 \, dx_4$ is an invariant. This invariant expression is the generalization of the orthogonal Euclidean or Galilean volume element. In Euclidean space, for example, with polar co-ordinates we have

$$g = \begin{vmatrix} 1, & 0, & 0, \\ 0, & r^2, & 0, \\ 0, & 0, & r^2 \sin^2 \theta \end{vmatrix},$$

(cf. **§ 36·2**). Hence

$$g^{1/2} = r^2 \sin \theta$$

and the invariant **(36·41)** becomes therefore

$$r^2 \sin \theta \, dr \, d\theta \, d\phi,$$

[1] Strictly this is only true in the integrated form:

$$\int J \, dx_1' \, dx_2' \, dx_3' \, dx_4' = \int dx_1 \, dx_2 \, dx_3 \, dx_4.$$

which is in fact the expression for the volume element which simple geometrical considerations yield.

§ 36·5. Parallel Displacement

Consider the vector (dx_1, dx_2, dx_3, dx_4) or (dx, dy, dz, dt). It represents a spacial shift (dx, dy, dz) of something or other, which occurs during the time interval, dt. The square of its absolute value is

$$ds^2 = g_{mn}\, dx_m\, dx_n, \quad . \quad . \quad . \quad . \quad (36\cdot5)$$

and we may describe ds as the **interval** between the points (x_1, x_2, x_3, x_4) and $(x_1 + dx_1, x_2 + dx_2, x_3 + dx_3, x_4 + dx_4)$. It is convenient to term the former point the origin of the vector (dx_1, dx_2, dx_3, dx_4). We are going to compare this vector and another one of the same absolute value, ds, whose origin is at a neighbouring point $(x_1 + \delta x_1, x_2 + \delta x_2, x_3 + \delta x_3, x_4 + \delta x_4)$. We shall assume that this second vector is related to the first by some law—it does not for the moment matter what law—so that we may represent the component corresponding to dx_m (or ξ^m as we shall sometimes call it) by $dx_m + \delta dx_m$, or by $\xi^m + \delta\xi^m$, since, for a given vector ξ^m, the new vector (which we may conveniently term the **displacement** of the original one, ξ^m) will depend on δx_m. In Fig. 36·5 the vector ξ^m is represented by oa and the displaced vector, $\xi^m + \delta\xi^m$, by bc, having been displaced by the amount ob or δx_m. The simplest kind of displacement will be one which is such that when the vector ob is displaced by the amount oa, i.e. when its origin is shifted from o to a, the end point, c, of the displaced vector will coincide with the end point of bc. This means that $\delta\, dx_m$ is identical with $d\, \delta x_m$. Since the opposite sides of the figure $oacb$ have equal lengths—the kind of displacement we are dealing with being one in which the absolute value of the displaced vector remains constant—it is appropriate to call the displacement a **parallel displacement.** In fact, oa and ob will be parallel in the ordinary sense of the term in a Euclidean or Galilean continuum.

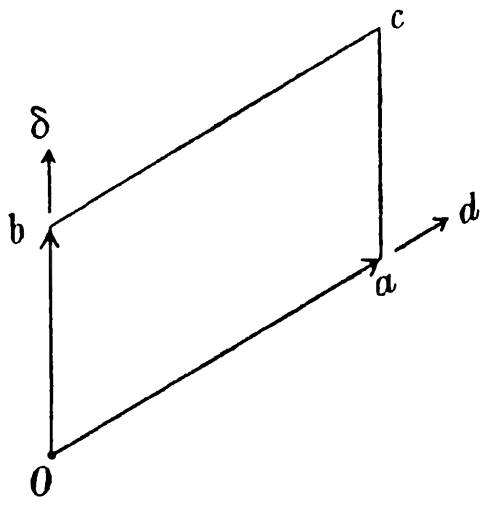

Fig. 36·5

We wish to find an expression for $\delta\xi^m$, the change in the value of ξ^m due to the displacement δx_m. We shall attempt to do this by writing

$$\delta\xi^m = A^m_{pq}\, \xi^p\, \delta x_q, \quad . \quad . \quad . \quad . \quad (36\cdot51)$$

the coefficients A^m_{pq} being independent of the components ξ^p and δx_q. Since ds, the absolute value of ξ^m, remains unaltered we must have

$$\delta(ds^2) \equiv \delta(g_{mn}\xi^m\xi^n) = 0,$$

or
$$\frac{\partial g_{mn}}{\partial x_r}\xi^m\xi^n\,\delta x_r + 2g_{mn}\xi^m\delta\xi^n = 0$$

and therefore

$$\frac{\partial g_{mn}}{\partial x_r}\xi^m\xi^n\,\delta x_r + 2g_{mn}\xi^m A^n_{pq}\xi^p\,\delta x_q = 0,$$

by **(36·51)**. We get, therefore, if we make suitable changes in the indices,

$$\left(\frac{\partial g_{mn}}{\partial x_r} + 2g_{no}\,A^o_{mr}\right)\xi^m\xi^n\,\delta x_r = 0.$$

This is true when $\xi^m\xi^n\,\delta x_r$ is arbitrary. Therefore

$$\frac{\partial g_{mn}}{\partial x_r} + 2g_{no}A^o_{mr} = 0. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(36·52)}$$

The equation **(36·51)** means

$$\delta\,dx_m = A^m_{pq}\,dx_p\,\delta x_q,$$

Similarly

$$d\,\delta x_m = A^m_{pq}\,\delta x_p\,dx_q.$$

Now $d\,\delta x_m = \delta\,dx_m$ and therefore

$$A^m_{pq} = A^m_{qp}. \quad . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(36·521)}$$

We have a large number of equations like **(36·52)**, one for each choice of m, n and r. Let us write down the two following, which have been derived from **(36·52)** by interchanging m and r, and n and r respectively:

$$\left.\begin{aligned}\frac{\partial g_{rn}}{\partial x_m} + 2g_{no}A^o_{mr} &= 0,\\ \frac{\partial g_{mr}}{\partial x_n} + 2g_{ro}A^o_{mn} &= 0.\end{aligned}\right\} \quad . \quad . \quad . \quad \textbf{(36·522)}$$

On adding the last two equations and subtracting **(36·52)** we get

$$\left(\frac{\partial g_{mr}}{\partial x_n} + \frac{\partial g_{nr}}{\partial x_m} - \frac{\partial g_{mn}}{\partial x_r}\right) + 2g_{ro}A^o_{mn} = 0,$$

or

$$\Gamma_{mn,\,r} + g_{ro}A^o_{mn} = 0, \quad . \quad . \quad . \quad \textbf{(36·523)}$$

by **(36·3)**.

On multiplying by g^{rs} and carrying out the implied summation we get

$$A^s_{mn} = -\Gamma^s_{mn}, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(36·53)}$$

by **(36·302)** and the formula **(36·51)** thus becomes

$$\delta\, dx_m + \Gamma^m_{pq}\, dx_p\, \delta x_q = 0 \quad . \quad . \quad . \quad \textbf{(36·54)}$$

and is therefore a generalization of **(36·301)**. We may regard it as a definition of parallelism in a Riemannian continuum.

It will be noticed, on comparing **(36·301)** and **(36·54)**, that **a geodesic is the result of parallel displacements of a vector dx_m in its own direction.**

§ 36·6. Covariant Derivation

A set of differential quotients like $\partial B_m/\partial x_n$ can only be regarded as a tensor while we use systems of orthogonal co-ordinates (cf. **§ 34·8**). It is very desirable and indeed essential that we should search for tensors whose components, when referred to orthogonal co-ordinate systems in Galilean space-time, become differential quotients of tensors. We shall start out from the invariant

$$d(B_m\, \delta x_m),$$

in which B_m is any covariant vector and d and δ have the same significance as in the last section. This invariant may be written

$$\frac{\partial B_m}{\partial x_n} dx_n\, \delta x_m + B_m d\, \delta x_m,$$

or

$$\frac{\partial B_m}{\partial x_n} dx_n\, \delta x_m - B_m \Gamma^m_{pq}\, dx_p\, \delta x_q,$$

by **(36·54)**. On making suitable changes in the indices we may put it in the form:

$$\left(\frac{\partial B_m}{\partial x_n} - B_o \Gamma^o_{mn}\right) dx_n\, \delta x_m.$$

As this is an invariant and $dx_n\, \delta x_m$ is an arbitrary tensor we conclude that

$$\frac{\partial B_m}{\partial x_n} - B_o \Gamma^o_{mn} \quad . \quad . \quad . \quad . \quad . \quad \textbf{(36·6)}$$

is a covariant tensor of rank two. It is called the **covariant derivative** of the vector B_m. On multiplying it by dx_n and carrying out the implied summation we get further:

$$dB_m - B_o \Gamma^o_{mn}\, dx_n \quad . \quad . \quad . \quad . \quad \textbf{(36·61)}$$

is a covariant vector. If we subtract from the tensor component

$$\frac{\partial B_m}{\partial x_n} - B_o \Gamma^o_{mn}$$

the component

$$\frac{\partial B_n}{\partial x_m} - B_o \Gamma^o_{nm}$$

we obtain, since $\Gamma^o_{nm} = \Gamma^o_{mn}$,

$$\frac{\partial B_m}{\partial x_n} - \frac{\partial B_n}{\partial x_m}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (36{\cdot}62)$$

which must be the mn component of a tensor. It represents the **curl** of **B**. It will be remembered that we have already anticipated (**§§ 35·2** and **36·1**) that the familiar expression in orthogonal co-ordinates for the component of a **curl** is valid in any system of co-ordinates.

Next consider the covariant vector

$$\frac{\partial}{\partial x_n}(A_m B^m).$$

It is a covariant vector because $A_m B^m$ is an invariant and the typical covariant vector is $\partial \Phi / \partial x_n$, where Φ is an invariant. If we write

$$\frac{\partial}{\partial x_n}(A_m B^m) = C_n$$

where C_n is a covariant vector, we have

$$\frac{\partial A_m}{\partial x_n} B^m + A_m \frac{\partial B^m}{\partial x_n} = C_n.$$

Now by **(36·6)**

$$\frac{\partial A_m}{\partial x_n} = T_{mn} + A_o \Gamma^o_{mn},$$

where T_{mn} is a certain covariant tensor. Therefore

$$T_{mn} B^m + A_o B^m \Gamma^o_{mn} + A_m \frac{\partial B^m}{\partial x_n} = C_n,$$

and so

$$A_o B^m \Gamma^o_{mn} + A_m \frac{\partial B^m}{\partial x_n} = D_n,$$

where D_n is the covariant vector $C_n - T_{mn} B^m$. Therefore

$$A_m \left(\frac{\partial B^m}{\partial x_n} + B^o \Gamma^m_{on} \right) = D_n,$$

and as A_m may be an arbitrary vector we conclude that

$$\left.\begin{aligned} &\frac{\partial B^m}{\partial x_n} + B^o \Gamma^m_{on}, \\ \text{and} \quad &dB^m + B^o \Gamma^m_{on}\, dx_n, \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (36{\cdot}63)$$

are tensors. The former of these is the **covariant derivative** of B^m.

A little difficulty seems to arise here. We have often assumed without a sufficiently close examination that the difference of two tensors of the same kind is a tensor, for example, that $C_n - T_{mn}B^m$ is a tensor. On the other hand, we have arrived at conclusions like **(36·61)** and **(36·63)** which make it clear that dB_m and dB^m, which, in fact, represent differences between vectors of the same kind, are not themselves vectors. The fact is, as we shall now show, that **the sum or difference of two tensors of the same kind is a tensor provided always they are tensors at the same place** (x_1, x_2, x_3, x_4). By **sum** or **difference** of two tensors is meant of course a tensor any one of whose components is the sum or difference of the corresponding components of the two tensors in question. Take, for example, two covariant vectors, A_m and B_m. The appropriate equations of transformation from one co-ordinate system to another are:

$$A_\mu' = \frac{\partial x_m}{\partial x_\mu'} A_m,$$

and

$$B_\mu' = \frac{\partial x_m}{\partial x_\mu'} B_m.$$

If A_m and B_m are vectors at the same place, (x_1, x_2, x_3, x_4), the differential quotients $\partial x_m/\partial x_\mu'$ have the same values for both vectors and hence

$$A_\mu' + B_\mu' = \frac{\partial x_m}{\partial x_\mu'}(A_m + B_m).$$

We understand now why it is quite correct to regard $C_n - T_{mn}B^m$ as a vector and that the reason why dB^m is not a vector is that it represents the difference of two vectors at *different places*.

When we apply the process of **contraction (§ 34·8)** to the first equation **(36·63)** we get

$$\frac{\partial B^m}{\partial x_m} + B^o\Gamma^m_{om}. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(36·64)}$$

This is the **divergence** of the vector B^m. It is appropriately so called because, in orthogonal co-ordinates in Galilean space-time, Γ^m_{om} vanishes, the $g_{\alpha\beta}$ being constants, and hence with such co-ordinates **(36·64)** reduces to

$$\partial B^m/\partial x_m \equiv \partial B^1/\partial x_1 + \partial B^2/\partial x_2 + \partial B^3/\partial x_3 + \partial B^4/\partial x_4.$$

It is easy to show that

$$\Gamma^m_{om} = \frac{1}{2} g^{mr} \frac{\partial g_{mr}}{\partial x_o}, \quad . \quad . \quad . \quad . \quad \textbf{(36·65)}$$

and since

$$\frac{1}{g}\frac{\partial g}{\partial x_o} = g^{mr}\frac{\partial g_{mr}}{\partial x_o}$$

we get

$$\Gamma^m_{om} = \frac{1}{g^{1/2}}\frac{\partial g^{1/2}}{\partial x_o}. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(36·66)}$$

The divergence of B^m may therefore be written

$$\frac{\partial B^m}{\partial x_m} + \frac{1}{g^{1/2}}\frac{\partial g^{1/2}}{\partial x_m}B^m,$$

or

$$\frac{1}{g^{1/2}}\left(g^{1/2}\frac{\partial B^m}{\partial x_m} + B^m\frac{\partial g^{1/2}}{\partial x_m}\right),$$

or finally

$$\frac{1}{g^{1/2}}\frac{\partial (g^{1/2}B^m)}{\partial x_m}. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(36·67)}$$

§ 36·7. Covariant Derivatives of Tensors of Higher Rank

Starting out from the statement

$$T_n{}^m A_m B^n \text{ is an invariant,}$$

$T_n{}^m$ being a mixed tensor and A_m and B^n being arbitrary covariant and contravariant vectors respectively, we have

$$\frac{\partial}{\partial x_r}(T_n{}^m A_m B^n) = C_r,$$

where C_r is a covariant vector. Therefore

$$A_m B^n\frac{\partial T_n{}^m}{\partial x_r} + B^n T_n{}^m\frac{\partial A_m}{\partial x_r} + A_m T_n{}^m\frac{\partial B^n}{\partial x_r} = C_r.$$

Now by **(36·6)** and **(36·63)**

$$\frac{\partial A_m}{\partial x_r} = \Gamma^o_{mr}A_o + \Phi_{mr},$$

and

$$\frac{\partial B^n}{\partial x_r} = -\Gamma^n_{or}B^o + \Psi_r{}^n,$$

where Φ_{mr} and $\Psi_r{}^n$ are certain tensors. Hence

$$A_m B^n\frac{\partial T_n{}^m}{\partial x_r} + B^n T_n{}^m(\Gamma^o_{mr}A_o + \Phi_{mr}) + A_m T_n{}^m(\Psi_r{}^n - \Gamma^n_{or}B^o) = C_r.$$

Therefore

$$A_m B^n\frac{\partial T_n{}^m}{\partial x_r} + A_o B^n T_n{}^m\Gamma^o_{mr} - A_m B^o T_n{}^m\Gamma^n_{or} = D_r,$$

where D_r means the covariant vector,

$$C_r - B^n T_n{}^m\Phi_{mr} - A_m T_n{}^m\Psi_r{}^n.$$

This result may be written:

$$A_m B^n \left\{ \frac{\partial T_n{}^m}{\partial x_r} + T_n{}^o \Gamma^m_{or} - T_o{}^m \Gamma^o_{nr} \right\} = D_r.$$

Now as A_m and B^n are arbitrary we infer (cf. § 34·8) that

$$\frac{\partial T_n{}^m}{\partial x_r} + \Gamma^m_{or} T_n{}^o - \Gamma^o_{nr} T_o{}^m \quad . \quad . \quad . \quad (36{\cdot}7)$$

is a tensor. It is the **covariant derivative** of $T_n{}^m$.

It is easy to show in a similar way that, for example,

$$\frac{\partial T^{mn}}{\partial x_r} + \Gamma^m_{or} T^{on} + \Gamma^n_{or} T^{mo} \quad . \quad . \quad . \quad (36{\cdot}71)$$

is a tensor: the **covariant derivative of** T^{mn}.

It is of some interest to study the divergence of T^{mn}. This will be

$$\frac{\partial T^{mn}}{\partial x_n} + \frac{1}{g^{1/2}} \frac{\partial g^{1/2}}{\partial x_o} T^{mo} + \Gamma^m_{on} T^{on}.$$

The Christoffel expression Γ^m_{on} is, as we have seen, symmetrical in o and n, and if it should happen that T^{on} is antisymmetrical in o and n, i.e. if $T^{on} = -T^{no}$, then the sum $\Gamma^m_{on} T^{on}$ vanishes identically; so that in this case the divergence of T^{mn} becomes

$$\boxed{\frac{1}{g^{1/2}} \frac{\partial (g^{1/2} T^{mn})}{\partial x_n}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (36{\cdot}72)$$

We might conclude this section by the following general statement which the reader will now have no difficulty in proving:

$$\frac{\partial T^{m\,n\,o\,\ldots}_{p\,q\,r\,\ldots}}{\partial x_k} + \Gamma^m_{\omega k} T^{\omega\,n\,o\,\ldots}_{p\,q\,r\,\ldots} + \Gamma^n_{\omega k} T^{m\,\omega\,o\,\ldots}_{p\,q\,r} + \Gamma^o_{\omega k} T^{m\,n\,\omega\,\ldots}_{p\,q\,r\,\ldots}$$
$$+ \ldots - \Gamma^\omega_{pk} T^{m\,n\,o\,\ldots}_{\omega\,q\,r\,\ldots} - \Gamma^\omega_{qk} T^{m\,n\,o\,\ldots}_{p\,\omega\,r\,\ldots} - \Gamma^\omega_{rk} T^{m\,n\,o\,\ldots}_{p\,q\,\omega\,\ldots}$$
$$- \ldots \text{ is a tensor.} \quad . \quad . \quad . \quad . \quad . \quad (36{\cdot}73)$$

It may easily be verified that the covariant derivatives of g_{mn} and g^{mn} are both of them identically zero. We have only to write down the covariant derivative of g_{mn} in order to recognize that it vanishes identically, thus

$$\frac{\partial g_{mn}}{\partial x_r} - \Gamma^\omega_{mr} g_{\omega n} - \Gamma^\omega_{nr} g_{m\omega}$$
$$= \frac{\partial g_{mn}}{\partial x_r} - \Gamma_{mr,\,n} - \Gamma_{nr,\,m}$$
$$= \frac{\partial g_{mn}}{\partial x_r} - \frac{1}{2}\left\{ \frac{\partial g_{mn}}{\partial x_r} + \frac{\partial g_{rn}}{\partial x_m} - \frac{\partial g_{mr}}{\partial x_n} \right\}$$
$$- \frac{1}{2}\left\{ \frac{\partial g_{nm}}{\partial x_r} + \frac{\partial g_{mr}}{\partial x_n} - \frac{\partial g_{nr}}{\partial x_m} \right\} = 0$$

identically. Hence

$$\frac{\partial g_{mn}}{\partial x_r} = \Gamma_{mr,\,n} + \Gamma_{nr,\,m} \quad . \quad . \quad . \quad (36{\cdot}74)$$

The covariant derivative of g^{mn} is

$$\frac{\partial g^{nk}}{\partial x_r} + g^{mk}\Gamma^n_{mr} + g^{no}\Gamma^k_{or}.$$

Let us represent it, for brevity, by X^{nk}_r. Therefore

$$\begin{aligned} g_{sk}g_{\tau n}X^{nk}_r &= g_{sk}g_{\tau n}\frac{\partial g^{nk}}{\partial x_r} + g_{sk}g^{mk}\Gamma_{mr,\,\tau} + g_{\tau n}g^{no}\Gamma_{or,\,s} \\ &= g_{sk}\left(-g^{nk}\frac{\partial g_{\tau n}}{\partial x_r}\right) + \Gamma_{sr,\,\tau} + \Gamma_{\tau r,\,s}, \\ &= -\frac{\partial g_{\tau s}}{\partial x_r} + \Gamma_{sr,\,\tau} + \Gamma_{\tau r,\,s}, \end{aligned}$$

which vanishes identically by **(36·74).**

§ 36·8. The Riemann-Christoffel Tensor

Writing the square of the interval, ds, in the form:

$$ds^2 = g_{mn}\xi^m\xi^n, \quad . \quad . \quad . \quad . \quad . \quad . \quad (36{\cdot}8)$$

and observing that in a parallel displacement of ds

$$\delta\xi^n = -\Gamma^n_{pq}\xi^p\,\delta x_q, \quad . \quad . \quad . \quad . \quad (36{\cdot}81)$$

we shall investigate the sum, $\oint\delta\xi^n$, obtained by adding up all

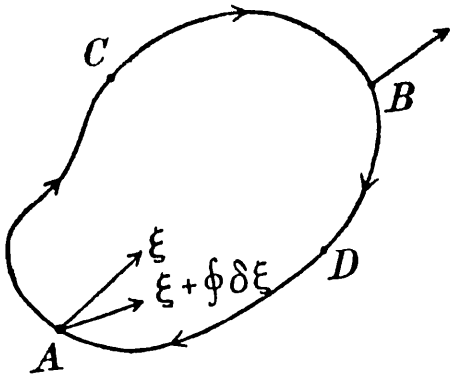

Fig. 36·8

the $\delta\xi^n$ resulting from a succession of parallel displacements carried out round a small closed loop, starting at any point, A (Fig. 36·8), and ending there. We have from **(36·81)**

$$\oint\delta\xi^n = -\oint\Gamma^n_{qp}\xi^p\,\delta x_q,$$

and therefore, on applying Stokes' theorem,

$$\oint \delta\xi^n = -\frac{1}{2}\iint\left\{\frac{\partial}{\partial x_p}(\Gamma^n_{qo}\xi^o) - \frac{\partial}{\partial x_q}(\Gamma^n_{po}\xi^o)\right\}dS^{pq},$$

in which the integration $\iint$ is extended over any surface bounded by the loop, and

$$dS^{pq} \equiv \delta x_p\, dx_q - \delta x_q\, dx_p$$

is a surface element and also a tensor. The bracket expression { } in the surface integral is equivalent to

$$\xi^o\left(\frac{\partial\Gamma^n_{qo}}{\partial x_p} - \frac{\partial\Gamma^n_{po}}{\partial x_q}\right) - \Gamma^n_{qo}\Gamma^o_{pw}\xi^w + \Gamma^n_{po}\Gamma^o_{qw}\xi^w,$$

or

$$\xi^o\left\{\frac{\partial\Gamma^n_{qo}}{\partial x_p} - \frac{\partial\Gamma^n_{po}}{\partial x_q} - \Gamma^n_{qw}\Gamma^w_{po} + \Gamma^n_{pw}\Gamma^w_{qo}\right\}. \quad . \quad \textbf{(36·82)}$$

Therefore

$$\oint \delta\xi^n = -\frac{1}{2}\iint\left\{\frac{\partial\Gamma^n_{qo}}{\partial x_p} - \frac{\partial\Gamma^n_{po}}{\partial x_q} - \Gamma^n_{qw}\Gamma^w_{po} + \Gamma^n_{pw}\Gamma^w_{qo}\right\}\xi^o dS^{pq}. \quad \textbf{(36·83)}$$

The integral $\int\delta\xi^n$ *is not* in general a vector; but obviously $\oint d\xi^n$ *is* a vector (cf. **§ 36·6**) since it means the difference of two vectors at the same place. If the loop round which $\oint$ is extended is a very small one the symbol, $\iint$, may be dropped and the surface integral becomes equal to minus the product of the bracket expression, { }, and 1/2 of $\xi^o dS^{pq}$. This product is a vector, namely $\oint d\xi^n$, and as $\xi^o dS^{pq}$ is an arbitrary tensor, we conclude that the bracket expression, { }, is a tensor. It is known as the **Riemann-Christoffel tensor** and we shall note it in the form :

$$\boxed{B^r_{mnk} \equiv \frac{\partial\Gamma^r_{km}}{\partial x_n} - \frac{\partial\Gamma^r_{mn}}{\partial x_k} - \Gamma^r_{ko}\Gamma^o_{mn} + \Gamma^r_{no}\Gamma^o_{km}.} \quad . \quad \textbf{(36·84)}$$

This is the most important tensor in Einstein's gravitational theory on account of the fact, established in **§ 36·9**, that its vanishing is the necessary and sufficient condition that the continuum shall be Galilean while at the same time it is derived solely from the g_{mn}. It has 4^4 or 256 components, of which,

however, only 20 are independent. We shall not deal in detail with this feature of the tensor; but will refer the reader to the works mentioned in the bibliography at the end of Chapter VII. Some of the relationships between the components are very obvious. For example

$$B^r_{mnk} + B^r_{kmn} + B^r_{nkm} = 0,$$

and

$$B^r_{mnk} = - B^r_{mkn}$$

§ 36·9. Flat and Curved Space-Time

These terms, ' flat ' and ' curved ', are not to be understood quite literally. We have already seen (cf. **§ 35·9**) that Galilean space-time, in which systems of co-ordinates can be so chosen that

$$ds^2 = \sum_n dx_n{}^2 \quad . \quad . \quad . \quad . \quad . \quad \textbf{(36·9)}$$

everywhere, resembles a Euclidean plane, and that the more comprehensive Riemannian continuum is exemplified by the surface of a sphere. We therefore sometimes speak of them as **flat** or **plane** space-time and **curved** space-time respectively. In flat space-time we can choose such co-ordinates that **(36·9)** is true everywhere and it is obvious that then every component of the Riemann-Christoffel tensor vanishes, i.e.

$$B^r_{mnk} = 0, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(36·91)}$$

or, as we shall briefly express it, the tensor vanishes, on account of the constancy of the g_{mn}. But the statement **(36·91)**, if true in one system of co-ordinates, must be true in any system of co-ordinates since

$$B'^{\rho}_{\mu\nu\kappa} = \frac{\partial x_\rho{}'}{\partial x_r} \frac{\partial x_m}{\partial x_\mu{}'} \frac{\partial x_n}{\partial x_\nu{}'} \frac{\partial x_k}{\partial x_\kappa{}'} B^r_{mnk} = 0$$

when **(36·91)** is true.

The converse proposition naturally suggests itself. Given that the Riemann-Christoffel tensor vanishes, or given that **(36·91)** is true for a continuum, can we infer that it is flat? In other words, can we infer that a co-ordinate system exists which gives the interval the form **(36·9)** everywhere, or, what amounts to the same thing, can we infer that a co-ordinate system exists in which the g_{mn} are everywhere constants? We shall see that this is the case. When **(36·91)** holds, the parallel displacement of a vector, **ds** ($\equiv dx_n$ or ξ^n), round a closed loop (Fig. 36·8), so that its origin starts from a point, A, and reaches this point again after travelling round the loop, will leave the

vector in coincidence with its original configuration, i.e. its terminal points will coincide with their original positions. This amounts to the same thing as to say that when the vector, **ds**, suffers a parallel displacement in which its origin travels along the loop from a point, A, to a point, B, the result will be independent of the direction taken, i.e. it will be all the same whether it travels along the loop by way of the point, C, or in the other direction, by way of the point, D.

We shall begin by studying a 2-dimensional continuum since it can be more easily illustrated by a figure. Let oa in Fig. 36·9 be a displacement vector, **ds**, and om a similar vector, **δs**;

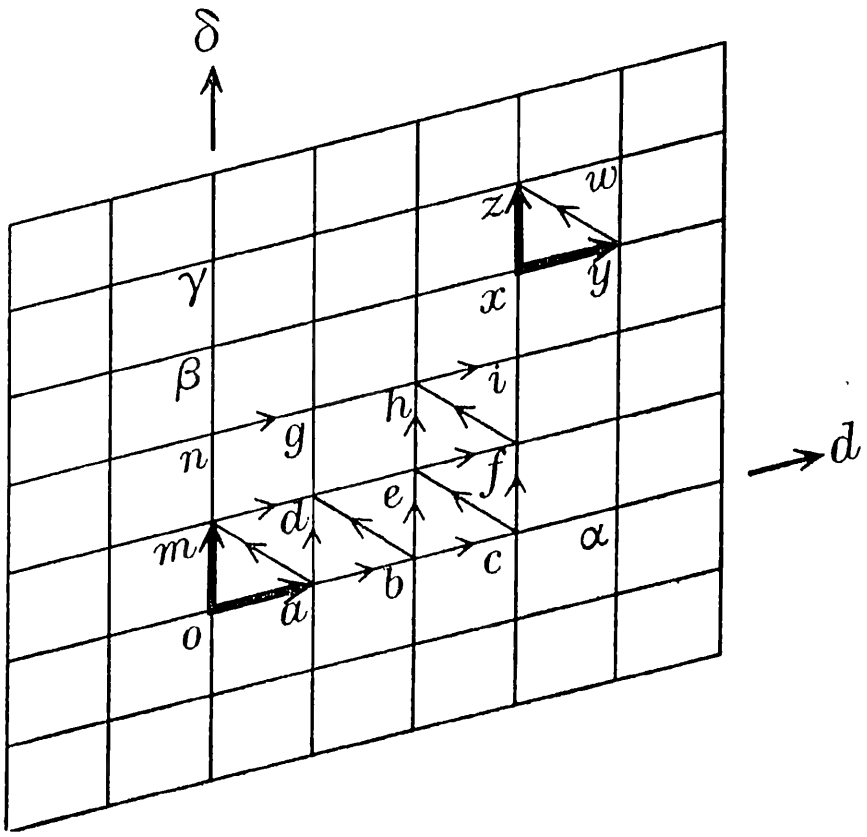

FIG. 36·9

both having the common origin, o. Now let oa suffer parallel displacement in its own direction, so that its origin occupies successively the positions a, b, c, etc., b being the terminal point of the vector when its origin is at a, c being its terminal point when its origin is at b, and so on. The geodesic $oabc\ldots$ is thus generated. Similarly we may generate another geodesic, $omn\ldots$, by the parallel displacement, in its own direction, of the vector **δs**. Next, let the vector, oa, suffer parallel displacements so that its origin travels along the geodesic $omn\ldots$ and occupies successively the points m, n, etc., the vectors md, ng, etc., being thus generated. In a similar way, by the parallel displacement of the vector, om, so that its origin, o, travels along the geodesic, $oabc\ldots$

and occupies successively the positions a, b, c . . . we shall produce the vectors ad, be, cf . . . If now all these vectors suffer parallel displacements, each in its own direction, the geodesics $mde'f'$ . . ., $ng'h'i'$ . . ., and adg' . . ., $be'h'$ . . ., $cf'i'$ . . . will be produced. The accented letters e', f', g', h' . . . denote points where the geodesics intersect and it should be observed that such points as g and g', for example, do not necessarily coincide. The former point terminates the vector ng, generated by the parallel displacement of oa so that its origin, o, is shifted along the geodesic omn . . . into the position n, whereas the latter point, g', is the intersection of the two geodesics generated by the parallel displacement of the vectors ng and ad each in its own direction. Certain points in the continuum have thus been defined or specified by certain parallel displacements of the vectors oa and om, namely, a, b, c . . . m, d, e, f . . . n, g . . . and other points, represented by accented letters, g', h', i', e', f', etc., which are the points of intersection of the two sets of geodesics described above. Fig. 36·9 represents a special case—namely, that where **(36·91)** holds—in which the points represented by accented letters, e.g. the points g', h', i', e', f' . . ., etc., coincide respectively with certain points, g, h, i, e, f . . ., etc., all of which have not yet been defined.

We can obviously found a system of co-ordinates on the two sets of geodesics. We may, for example, adopt the distances oc or $\int_o^c \mathbf{ds}$ and on or $\int_o^n \boldsymbol{\delta s}$ as the co-ordinates of the point, i', the intersection of the two geodesics $cf'i'$ . . . and $ng'h'i'$ . . . respectively and denote them by x_1 and x_2. If the geodesics are very near together, so that the meshes which they form are very fine, we may denote the differences between the co-ordinates of the points e' and f', for example, and between those of e' and h', i.e. the distances bc and mn, by dx_1 and dx_2 respectively, and the distance $f'h'$ by ds. We shall then have of course

$$ds^2 = g_{11}dx_1^2 + 2g_{12}\,dx_1\,dx_2 + g_{22}\,dx_2^2,$$

the coefficients g_{11}, g_{12}, g_{22} being in general functions of x_1 and x_2.

When the condition **(36·91)** is satisfied we may define a set of points o, a, b, c . . . m, d, e, f . . . n, g . . . , just as before and in addition other points h, i . . . x, y . . . which will, in consequence of **(36·91)**, coincide with the points already defined as the points of intersection of two sets of geodesics and denoted by accented letters g', h', e', f' . . . The point g, for example, has already been defined as the terminal point of the vector ng which is the result of that parallel displacement of the vector oa in which its origin, o, has

travelled along the geodesic o, m, n . . . to the point n. Now the vector dg is in any case, i.e. whether **(36·91)** is satisfied or not, the parallel displacement of mn when its origin is displaced from m to d, since ng is the parallel displacement of md. But when **(36·91)** is true dg will also result from the parallel displacement of om when its origin travels along *any* path to d and hence when its origin, o, is shifted first to a and then to d. Thus when **(36·91)** holds, the point g coincides with g', the intersection of the two geodesics adg' . . . and $ngh'i'$ . . . Similarly we can define a point h as the point where a arrives in consequence of the parallel displacement of oa in which its origin, o, travels along any path from o to g, provided **(36·91)** is true; or equally well as the point where m arrives in consequence of the parallel displacement of om in which its origin, o, travels from o to e. All the points defined in this way and represented by unaccented letters coincide with the corresponding points of intersection of two sets of geodesics which have been represented by the same letters accented.

Now it is a consequence of **(36·91)** that md is the parallel displacement of ab. Hence bd, ce, fh . . . yz . . . can be produced by the parallel displacement of am. Therefore the length yz, for example, is equal to the length am and we may call each of them ds. The difference of the co-ordinates of x and z is the distance $\beta\gamma$. And this is equal to the distances om and xz since all of them can be generated one from another by parallel displacement. Similarly the difference of the co-ordinates of x and y is equal to the distance xy and also to the distance oa. We have then

$$xy = oa = dx_1,$$
$$xz = om = dx_2,$$

and

$$yz = am = ds.$$

Now

$$ds^2 = g_{11}\,dx_1^2 + 2g_{12}\,dx_1\,dx_2 + g_{22}\,dx_2^2, \text{ (triangle } oam)$$

and

$$ds^2 = g_{11}'\,dx_1^2 + 2g_{12}'\,dx_1\,dx_2 + g_{22}'\,dx_2^2, \text{ (triangle } xyz)$$

and since dx_1 and dx_2 are quite arbitrary it follows that

$$g_{11} = g_{11}',$$
$$g_{12} = g_{12}',$$
$$g_{22} = g_{22}'.$$

That is to say the values of the coefficients g_{mn} are independent of the position of the triangle or mesh and they are therefore constants and indeed with our choice of co-ordinates g_{11} and g_{22} are each equal to unity. This follows because the respective

co-ordinate differences dx_1 and dx_2 between the points o and a and between the points o and m are equal to the actual distances oa and om; and since oa and om are arbitrary, we have in the special case where oa vanishes

$$am = om,$$

or $$ds^2 = dx_2{}^2;$$

but in this case $$ds^2 = g_{22}\, dx_2{}^2.$$

Hence $$g_{22} = 1.$$

We have learned that, in consequence of **(36·91)** the g_{mn} are constants. The 2-dimensional continuum is therefore Euclidean or, as we sometimes say, flat. The argument we have used can obviously be extended to a continuum of 3, 4 or more dimensions. We have thus established that **(36·91),** or the vanishing of the Riemann-Christoffel tensor, is a sufficient condition for flat space-time and we proved it to be a necessary condition at the outset of this section.

CHAPTER VI

THE GENERAL THEORY OF RELATIVITY

§ 37. Einstein's Law of Gravitation

WE have seen **(§ 36)** that the motion of a particle in a gravitational field is along a space-time geodesic and the law of gravitation will therefore be expressed by equations which determine for us the components, g_{mn}, of the fundamental tensor in any system of co-ordinates. This amounts **(§ 36·1)** to imposing some sort of limitation on the character of the Riemannian continuum. We have just had an example of such a thing, namely, that expressed by the vanishing of the Riemann-Christoffel tensor, or

$$B^r_{mnk} = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(37)}$$

This condition causes all the geodesics to be straight lines and banishes gravitational fields (in the strict or old-fashioned sense of the term) altogether. We have therefore to search for some less stringent statement than **(37)**—one which will assign to the geodesics such forms that the motion of particles along them will be in accord with our experience; and whatever the statement may be, it must include **(37)** as a special case, or at any rate it must be capable of approaching it very closely, since we are bound to provide for the case where gravitational forces vanish or become very small. Einstein's original choice for the **law of gravitation** (in empty space with neighbouring gravitating masses) was the statement:

$$G_{mn} \equiv B^r_{mnr} \equiv \frac{\partial \Gamma^r_{rm}}{\partial x_n} - \frac{\partial \Gamma^r_{mn}}{\partial x_r} - \Gamma^r_{ro}\Gamma^o_{mn} + \Gamma^r_{no}\Gamma^o_{mr} = 0. \qquad \textbf{(37·01)}$$

This choice is consistent with condition just mentioned; since when **(37)** is true **(37·01)** necessarily follows. This is because the components, G_{mn}, are sums of certain components of the Riemann-Christoffel tensor. In fact, the particular component G_{mn} is equal to $B^1_{mn1} + B^2_{mn2} + B^3_{mn3} + B^4_{mn4}$. On the other hand **(37·01)** may be true even when **(37)** does not hold.[1]

[1] Since in order that $a + b + c$ may vanish it is not necessary that a, b and c vanish separately.

The tensor $G_{mn} \equiv B^r_{mnr}$ is not, at first sight, the only one which we derive from B^r_{mnk} by contraction. The two further possibilities

$$B^r_{mrk} \text{ and } B^r_{rnk}$$

have to be considered. Of these two the former is identical with the tensor G_{mn} except as regards its sign; since

$$B^r_{mrk} = -G_{mk},$$

and therefore its adoption would give the same law as **(37·01)**. The latter tensor, B^r_{rnk}, is

$$B^r_{rnk} \equiv \frac{\partial \Gamma^r_{kr}}{\partial x_n} - \frac{\partial \Gamma^r_{nr}}{\partial x_k}, \quad \ldots\ldots\ldots \textbf{(37·011)}$$

$$= \frac{\partial^2}{\partial x_n \, \partial x_k}(\log g^{1/2}) - \frac{\partial^2}{\partial x_k \, \partial x_n}(\log g^{1/2})$$

and vanishes identically. It is therefore useless to us.

The tensor G_{mn} is consequently all we have—unless indeed we develop fundamental tensors from B^r_{mnk} containing differential quotients of the g_{mn} of higher order than the second—for expressing the law of gravitation.

Equation **(37·01)** expresses Einstein's law of gravitation for empty regions—regions free from masses or electromagnetic fields—and, as we shall find when we investigate it further, it is in extremely close agreement with Newton's law. The small difference between the two is an actual gain, since it accounts within the limits of observational error for the discrepancy between the actual motion of the planet Mercury and that required by Newton's law.

Furthermore, Einstein's law predicts, as he himself showed, two new phenomena, both of them observed later and found to be in quantitative agreement with the theory within the limits of observational error: (i) a refraction of the rays of light passing a gravitating mass such as a star—as if the region round the star were a converging lens of low power—and (ii) a displacement of the spectrum lines towards the red end of the spectrum in the case of the light from a star—or, more precisely, from a place where the gravitational potential is very much lower than that of the observer (cf. **§§ 37·4** and **37·5**).

There is another suggestion for the law of gravitation which makes use of the tensor G_{mn}.[1] We may identify this tensor

[1] This law **(37·02)** does not appear to have been adopted by Einstein for empty regions; but his 'Kosmologische Betrachtungen', published in the *Sitzungsberichte der Preussischen Akademie der Wissenschaften*, 1917, strongly suggest it.

with g_{mn}—apart from a constant factor—so that, instead of **(37·01)**, we have

$$G_{mn} = \lambda g_{mn} . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(37·02)}$$

Since **(37·01)** is in excellent agreement with the important observations already mentioned it is clear that the constant, λ, which is known as the **cosmic constant,** must be a very small quantity (cf. **§ 38·6**).

The law **(37·02)** has one advantage over **(37·01).** It predicts that remote material configurations (e.g. the spiral nebulae) must be receding from us—if we disregard exceptional cases where the recession is compensated by gravitational attraction—and also a slowing down of atomic vibrations. Both of these theoretical consequences would manifest themselves to observation as a displacement of the spectral lines, in the light emitted from the configurations in question, towards the red end. Such a displacement is actually observed in the case of most of the spiral nebulae **(§ 38·6).**

§ 37·1. Gravitational Field of an Isolated Particle

The importance of this problem—from the point of view of the theory of relativity—is due mainly to the fact that we already know its solution to a first and quite close approximation. In approaching it we shall make use of the observed fact that the gravitational field is static, that is to say the g_{mn} are not functions of the time; and of the symmetry which 'isolated particle' implies. The g_{mn} must therefore be functions of r only and in the expression for ds^2 all terms involving the products $dr\,d\theta$, $dr\,d\phi$, and $d\theta\,d\phi$ must vanish and finally the terms containing the products $dr\,dt$, $d\theta\,dt$, and $d\phi\,dt$ must also vanish. We are thus forced to the expression

$$ds^2 = f_1(r)dr^2 + f_2(r)r^2\,d\theta^2 + f_3(r)r^2\sin^2\theta\,d\phi^2 - f_4(r)c^2\,dt^2 \quad . \quad \textbf{(37·1)}$$

in which obviously the values of the functions, f, approach very closely to unity. It is also fairly obvious that the symmetry requires that the functions $f_2(r)$ and $f_3(r)$ must be identical. Consider, for example, the case where the spacial part of ds lies in the surface of the sphere of radius r (with $r = 0$ as centre). In this case

$$ds^2 = f_2(r)r^2\,d\theta^2 + f_3(r)r^2\sin^2\theta\,d\phi^2 - f_4(r)c^2\,dt^2.$$

A special case of this is

$$ds^2 = f_3(r)r^2\,d\phi^2 - f_4(r)c^2\,dt^2$$

and we have already to rotate the spacial reference system suitably to make this become

$$ds^2 = f_3(r)r^2\,d\theta^2 - f_4(r)c^2\,dt^2.$$

Thus we may identify f_2 and f_3. Let us now introduce a new co-ordinate, r_1, defined by

$$r^2 f_2(r) = r^2 f_3(r) = r_1^2,$$

so that we may now write

$$ds^2 = \psi_1(r_1)dr_1^2 + r_1^2\,d\theta^2 + r_1^2\sin^2\theta\,d\phi^2 - \psi_2(r_1)c^2\,dt^2. \quad . \quad \textbf{(37·11)}$$

The radial co-ordinate, r_1, does not, of course, represent radial distances as measured by a rod, since a measuring rod extending from r_1 to $r_1 + dr_1$ ($d\theta = d\phi = dt = 0$) has the length

$$\sqrt{ds^2} = \sqrt{\psi_1(r_1)}\,dr_1.$$

In what follows we shall drop the subscript and write r for r_1. It is advantageous and usual to express ds^2 in the form

$$ds^2 = e^L dr^2 + r^2\,d\theta^2 + r^2\sin^2\theta\,d\phi^2 - e^N c^2\,dt^2, \quad . \quad \textbf{(37·12)}$$

L and N being respectively the natural logarithms of ψ_1 and ψ_2. Hence

$$\left.\begin{aligned} g_{11} &= e^L,\ g_{22} = r^2 \ (\text{or } x_1^2),\\ g_{33} &= r^2\sin^2\theta \ (\text{or } x_1^2\sin^2 x_2),\\ g_{44} &= -e^N c^2, \end{aligned}\right\} \quad . \quad . \quad \textbf{(37·121)}$$

the remaining g_{mn} vanishing. We easily obtain for the determinant, g, and its square root the expressions:

$$\left.\begin{aligned} g &= -c^2 r^4\sin^2\theta e^{L+N},\\ g^{1/2} &= \sqrt{-1}\,cr^2\sin\theta e^{\frac{1}{2}(L+N)}, \end{aligned}\right\} \quad . \quad . \quad \textbf{(37·122)}$$

respectively.

Turning to the law of gravitation **(37·01)** we note that it may be written

$$G_{mn} \equiv \frac{\partial^2(\log g^{1/2})}{\partial x_n\,dx_m} - \frac{\partial \Gamma^r_{mn}}{\partial x_r} - \Gamma^o_{mn}\frac{\partial}{\partial x_o}(\log g^{1/2}) + \Gamma^r_{no}\Gamma^o_{rm} = 0. \quad . \quad . \quad \textbf{(37·13)}$$

In his original paper on the general theory of relativity Einstein proceeded in a different way and chose co-ordinates for which g was constant and equal to unity. He thus simplified the expression of the law of gravitation to

$$\frac{\partial \Gamma^r_{mn}}{\partial x_r} - \Gamma^r_{no}\Gamma^o_{rm} = 0. \quad . \quad . \quad . \quad \textbf{(37·131)}$$

The symmetries in our problem have already partly determined the g_{mn} and all that remains to be done is the evaluation of the two functions of r, namely e^L and e^N. This we can do by sub-

stituting the expressions **(37·121)** for the g_{mn} first of all in the various components of Γ^p_{qr} and then finally in **(37·13)**. There are 40 components, Γ^p_{qr}, in the most general case. For any value of p, say $p = 2$, there are $4^2 = 16$ possibilities for qr; but the symmetry ($\Gamma^r_{qp} = \Gamma^r_{pq}$) reduces this number to 10. Hence, when we remember the 4 possibilities for p, we get 40 components in all. Γ^p_{qr} is not, of course, a tensor.

It is easily seen that, with the values of g_{mn} given above,

$$\Gamma^p_{qr} = 0 \text{ when } p \neq q \neq r,$$

and of the other components

$$\left.\begin{aligned} \Gamma^p_{pp} &= \frac{\partial}{dx_p}(\log g^{1/2}_{pp}), \\ \Gamma^p_{qq} &= -\frac{1}{2g_{pp}}\frac{\partial g_{qq}}{dx_p}, \\ \Gamma^p_{pq} &= \frac{1}{2g_{pp}}\frac{\partial g_{pp}}{\partial x_q}. \end{aligned}\right\} \textit{No summation is implied and } p \neq q. \qquad \textbf{(37·14)}$$

When we examine all the possibilities under **(37·14)** we find that they reduce to

$$\left.\begin{aligned} &\Gamma^1_{11} = dL/2dr,\ \Gamma^1_{22} = -re^{-L}, \\ &\Gamma^1_{33} = -e^{-L}r\sin^2\theta,\ \Gamma^1_{44} = c^2e^{N-L}dN/2dr, \\ &\Gamma^2_{33} = -\sin\theta\cos\theta,\ \Gamma^2_{12} = 1/r,\ \Gamma^3_{13} = 1/r, \\ &\Gamma^3_{23} = \cot\theta,\ \Gamma^4_{14} = dN/2\,dr. \end{aligned}\right\}. \qquad \textbf{(37·15)}$$

These expressions have to be substituted in the G_{mn} of **(37·13)**. We shall work out

$$G_{11} = 0$$

as an illustration and leave the rest to the reader.

The first term in **(37·13)** is

$$\frac{\partial^2}{\partial x_n\,\partial x_m}(\log g^{1/2}) = \frac{\partial}{\partial r}\left\{\frac{1}{2g}\frac{\partial g}{\partial r}\right\} = -\frac{2}{r^2} + \frac{1}{2}(L'' + N'') \qquad \textbf{(37·16)}$$

where $L'' \equiv d^2L/dr^2$ and $N'' \equiv d^2N/dr^2$.

The next term is

$$\begin{aligned} -\frac{\partial\Gamma^r_{11}}{\partial x_r} &= -\left\{\frac{\partial\Gamma^1_{11}}{\partial r} + \frac{\partial\Gamma^2_{11}}{\partial\theta} + \frac{\partial\Gamma^3_{11}}{\partial\phi} + \frac{\partial\Gamma^4_{11}}{\partial t}\right\} \\ &= -\frac{\partial\Gamma^1_{11}}{\partial r} \\ &= -L''/2. \qquad \textbf{(37·161)} \end{aligned}$$

The third term is

$$-\Gamma^0_{11}\frac{\partial}{\partial x_o}(\log g^{1/2}) = -\Gamma^1_{11}\frac{\partial}{dr}(\log g^{1/2})$$

$$-\Gamma^2_{11}\frac{\partial}{\partial\theta}(\log g^{1/2}) - \Gamma^3_{11}\frac{\partial}{\partial\phi}(\log g^{1/2}) - \Gamma^4_{11}\frac{\partial}{dt}(\log g^{1/2})$$

$$= -\Gamma^1_{11}\frac{\partial}{dr}(\log g^{1/2}),$$

$$= -\frac{1}{2}L'\left\{\frac{1}{2g}\frac{\partial g}{\partial r}\right\},$$

$$= -\frac{1}{2}L'\left\{\frac{2}{r}+\frac{L'+N'}{2}\right\}$$

$$= -\frac{L'}{r}-\frac{L'^2}{4}-\frac{L'N'}{4}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(37·162)}$$

where $L' \equiv dL/dr$ and $N' \equiv dN/dr$.

The last term of **(37·13)** is the sum of 16 terms of which only the following survive:

$$(\Gamma^1_{11})^2 + (\Gamma^2_{12})^2 + (\Gamma^3_{13})^2 + (\Gamma^4_{14})^2$$

$$= \left(\frac{1}{2}L'\right)^2 + \frac{1}{r^2} + \frac{1}{r^2} + \frac{N'^2}{4}$$

$$= \frac{L'^2}{4} + \frac{N'^2}{4} + \frac{2}{r^2}. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(37·163)}$$

On adding together the contributions **(37·16)** **(37·161)** **(37·162)** and **(37·163)** we obtain

$$\left.\begin{aligned} &G_{11} \equiv \frac{1}{2}N'' - \frac{L'N'}{4} + \frac{N'^2}{4} - \frac{L'}{r} = 0, \\ &\text{and to this we may add:} \\ &G_{22} \equiv -e^{-L}\left\{r\left(\frac{L'-N'}{2}\right)-1\right\} - 1 = 0, \\ &G_{33} \equiv -\sin^2\theta\,.\left[e^{-L}\left\{r\left(\frac{L'-N'}{2}\right)-1\right\}+1\right] = 0, \\ &G_{44} \equiv -c^2e^{N-L}\left\{\frac{N''}{2}-\frac{N'L'}{4}+\frac{N'^2}{4}+\frac{N'}{r}\right\} = 0, \end{aligned}\right\} \quad \textbf{(37·17)}$$

which the reader may work out for himself. Of the remaining G_{mn} they either vanish identically or the terms in them vanish

individually. It is easy to deal with these equations. The first and the fourth yield at once

$$-\frac{L'}{r} = \frac{N'}{r},$$

or

$$L' + N' = 0,$$

and consequently also

$$L + N = 0, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(37·18)}$$

since the constant of integration must vanish, on account of the fact that both L and N must approach zero as r approaches ∞, if the geometry is to be Galilean in empty space at great distances from gravitating masses.

The second and third of the equations **(37·17)** amount to the same thing. Let us substitute for L and L' in the second equation, using **(37·18)**. We thus obtain

$$e^N\left\{r\frac{dN}{dr} + 1\right\} - 1 = 0,$$

or

$$\frac{d}{dr}(re^N) = 1,$$

therefore

$$re^N = r + A,$$

or

$$e^N = 1 + \frac{A}{r}, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(37·19)}$$

where A is a constant of integration the significance of which we shall investigate later.

We are only concerned with the cases where A/r is very small compared with unity and it follows from **(37·18)** and **(37·19)** that

$$e^L = e^{-N} = 1/(1 + A/r) = 1 - A/r \quad . \quad \textbf{(37·191)}$$

to a sufficiently good approximation for all or most of our purposes.

§ 37·2. Planetary Motion

The isolated particle of the preceding section now becomes the sun and the investigation of planetary motion becomes that of the geodesics in a space-time such as we studied in the preceding section. We are of course dealing with the ideal case of a single planet whose mass is so small compared with that of the sun that the character of the space is practically determined by the latter. Our immediate task is very straightforward. All we have to do is to make appropriate substitutions in the geodesic equations

$$\frac{d^2x_i}{ds^2} + \Gamma^i_{mn}\frac{dx_m}{ds}\frac{dx_n}{ds} = 0. \quad . \quad . \quad . \quad . \quad \textbf{(37·2)}$$

Let us begin with $x_i = x_2 = 0$, and substitute for the Γ^2_{mn} the expressions in **(37·15)**. We thus obtain

$$\frac{d^2x_2}{ds^2} + \Gamma^2_{12}\frac{dx_1}{ds}\frac{dx_2}{ds} + \Gamma^2_{21}\frac{dx_2}{ds}\frac{dx_1}{ds} + \Gamma^2_{33}\frac{dx_3}{ds}\frac{dx_3}{ds} = 0,$$

or

$$\left.\begin{aligned} \frac{d^2\theta}{ds^2} + \frac{2}{r}\frac{d\theta}{ds}\frac{dr}{ds} - \sin\theta\cos\theta\left(\frac{d\phi}{ds}\right)^2 &= 0, \\ \text{and similarly} \\ \frac{d^2\phi}{ds^2} + \frac{2}{r}\frac{dr}{ds}\frac{d\phi}{ds} + 2\cot\theta\frac{d\theta}{ds}\frac{d\phi}{ds} &= 0, \\ \frac{d^2t}{ds^2} + N'\frac{dr}{ds}\frac{dt}{ds} &= 0. \end{aligned}\right\} \quad \textbf{(37·21)}$$

We shall not need the equation containing d^2r/ds^2. We shall use instead **(37·12)** which will play the part of the energy equation of classical dynamics (cf. **§ 8·5**). We may write it in the form:

$$1 = S^{-1}\left(\frac{dr}{ds}\right)^2 + r^2\left(\frac{d\theta}{ds}\right)^2 + r^2\sin^2\theta\left(\frac{d\phi}{ds}\right)^2 - Sc^2\left(\frac{dt}{ds}\right)^2, \quad \textbf{(37·211)}$$

in which e^N and e^L have been represented by S and S^{-1}, so that

$$S \equiv 1 + \frac{A}{r} \quad . \quad . \quad . \quad . \quad . \quad \textbf{(37·22)}$$

Let us so choose our reference system (r, θ, ϕ, t) that initially $d\theta/ds$ is zero and $\theta = \pi/2$. We then find from **(37·21)** that

$$\frac{d^2\theta}{ds^2} = 0,$$

so that

$$\frac{d\theta}{ds} = \text{constant};$$

and as the initial value of $d\theta/ds$ is zero the constant in question must be zero. Therefore $d\theta/ds$ is always zero and the motion of the planetary body is confined to the plane $\theta = \pi/2$.

Turning to the second of the equations **(37·21)**, we represent $d\phi/ds$ by α. Hence

$$\frac{d\alpha}{ds} + \frac{2}{r}\alpha\frac{dr}{ds} = 0,$$

or

$$\frac{d\alpha}{\alpha} + 2\frac{dr}{r} = 0.$$

Therefore

$$r^2\alpha = \text{constant}$$

or

$$r^2\frac{d\phi}{ds} = H \quad . \quad . \quad . \quad . \quad . \quad \textbf{(37·23)}$$

where H is some constant.

We can deal equally easily with the third equation **(37·21)**. It is convenient to represent dt/ds by β. Thus

$$\frac{d\beta}{ds} + \frac{dN}{ds}\beta = 0,$$

and therefore

$$\frac{d\beta}{\beta} + dN = 0.$$

Consequently

$$\beta = \text{constant} \times e^{-N},$$

or

$$\frac{dt}{ds} = \frac{K}{S} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(37·24)}$$

where K is a new constant and $S = e^N = 1 + A/r$.

We now turn to **(37·211)** which, since $d\theta/ds = 0$ and $\sin\theta = 1$, simplifies to

$$1 = S^{-1}\left(\frac{dr}{ds}\right)^2 + r^2\left(\frac{d\phi}{ds}\right)^2 - \frac{c^2K^2}{S},$$

when we substitute K/S for dt/ds in virtue of **(37·24)**. On making use of **(37·22)** and **(37·23)** we get

$$\left(\frac{dr}{ds}\right)^2 + r^2\left(\frac{d\phi}{ds}\right)^2 = S + c^2K^2 - \frac{AH^2}{r^3}. \quad . \quad . \quad \textbf{(37·25)}$$

Now let us get rid of ds by using **(37·23)**. We thus obtain

$$\left(\frac{dr}{d\phi}\right)^2 + r^2 = \frac{S + c^2K^2}{H^2}r^4 - Ar.$$

Lastly, adopting a well-known device, we introduce a new variable, u, defined by $u = 1/r$, as in **§ 5·5**, and obtain

$$\left(\frac{du}{d\phi}\right)^2 + u^2 = \frac{S + c^2K^2}{H^2} - Au^3,$$

or

$$\left(\frac{du}{d\phi}\right)^2 + u^2 = \frac{1 + c^2K^2}{H^2} + \frac{A}{H^2}u - Au^3,$$

since

$$S = 1 + A/r = 1 + Au.$$

On differentiating and dividing through by $2du/d\phi$ we get finally

$$\boxed{\frac{d^2u}{d\phi^2} + u = \frac{A}{2H^2} - \frac{3A}{2}u^2.} \quad . \quad . \quad . \quad \textbf{(37·26)}$$

The corresponding Newtonian equation (cf. **§ 5·5**) is

$$\frac{d^2u}{d\phi^2} + u = \frac{\kappa M}{h^2}, \quad . \quad . \quad . \quad . \quad \textbf{(37·261)}$$

where κ is the constant of gravitation, M is the mass of the sum and h is $r^2\,d\phi/dt$, which in the Newtonian theory is a constant.

The Newtonian theory of planetary motion is very nearly correct and we surmise that, whatever the significance of the constant, A, may turn out to be, the term $3Au^2/2$ in **(37·26)** must be a very small one, while $A/2H^2$ must approximate closely to $\kappa M/h^2$.

Let us then write

$$\frac{A}{2H^2} = \frac{\kappa M}{h^2}, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(37·262)}$$

and make use of the approximate relationship

$$ds^2 = -c^2\,dt^2, \quad . \quad . \quad . \quad . \quad \textbf{(37·263)}$$

which is near the truth in the case of planetary motion, since planetary velocities are very small compared with c. These two equations yield, when we take **(37·23)** into account,

$$A = -2\kappa M/c^2 \quad . \quad . \quad . \quad . \quad . \quad \textbf{(37·27)}$$

We have thus discovered the significance of the constant, A. Apart from the factor, $-2\kappa/c^2$, it must be identified with the mass of the sun or of the isolated particle of the preceding section, and the additional term, $-3Au^2/2$, of **(37·26)** turns out to be very small compared with $A/2H^2$.

We shall now investigate the effect of this minute additional term, $-3Au^2/2$. The general solution of the equation which we obtain when we ignore it may be written:

$$u = \frac{A}{2H^2}\{1 + e\cos(\phi - \eta)\}, \quad . \quad . \quad \textbf{(37·271)}$$

where e (the eccentricity of the orbit) and η are constants of integration. It will suffice for all our purposes to substitute this approximate expression for u in the small term, $-3Au^2/2$. The equation **(37·26)** will then become

$$\frac{d^2u}{d\phi^2} + u = \frac{A}{2H^2} - \frac{3A^3}{8H^4} - \frac{3A^3}{4H^4}e\cos(\phi - \eta) - \frac{3A^3}{8H^4}e^2\cos^2(\phi - \eta). \quad . \quad . \quad \textbf{(37·272)}$$

All the terms on the right except $A/2H^2$ are very small and of the same order of magnitude. The second and fourth may be ignored; but not the third one, containing the factor $\cos(\phi - \eta)$, notwithstanding the fact that it is of the same order of magnitude as those ignored. The reason for this is that it plays the same part in the equation as does the expression for a periodic impressed force in the equation of a simple harmonic system in which dissipative forces are absent and in which the natural period and that of the impressed force are equal (resonance).

The effect of this small term is therefore cumulative. Let us consider it by itself. That is to say, let us find a particular solution of

$$\frac{d^2u}{d\phi^2} + u = P\cos(\phi - \eta), \quad . \quad . \quad . \quad . \quad \textbf{(37·28)}$$

where P has been written for $-3A^3e/4H^4$. Such a solution can be put in the form

$$u = Q\phi\cos(\phi - \eta - \omega),$$

Q and ω being certain constants. On substituting in **(37·28)** we get

$$P\cos(\phi - \eta) = -2Q\sin(\phi - \eta - \omega),$$

or $$P\cos(\phi - \eta) = 2Q\{\cos(\phi - \eta)\sin\omega - \sin(\phi - \eta)\cos\omega\}.$$

Hence $$\sin\omega = 1,$$

and $$Q = P/2,$$

so that the particular solution in question is

$$u = \frac{P}{2}\phi\sin(\phi - \eta)$$

and the effect of the small term (in **37·272**) containing $\cos(\phi - \eta)$ is a contribution to u which is proportional to ϕ and which therefore gets greater and greater as ϕ increases, until at last it has accumulated to an observable quantity.

The approximate Newtonian solution **(37·271)** has to be modified therefore to

$$u = \frac{A}{2H^2}\{1 + e\cos(\phi - \eta)\} - \frac{3A^3}{8H^4}e\phi\sin(\phi - \eta),$$

or $$u = \frac{A}{2H^2}\{1 + e\cos(\phi - \eta) - \frac{3A^2}{4H^2}e\phi\sin(\phi - \eta)\};$$

which may be written

$$u = \frac{A}{2H^2}\{1 + eR(\cos\psi\cos(\phi - \eta) - \sin\psi\sin(\phi - \eta))\},$$

where $$R \equiv \sqrt{\left(1 + \frac{9A^4\phi^2}{16H^4}\right)}.$$

R is practically unity; therefore

$$u = \frac{A}{2H^2}\{1 + e[\cos(\phi - \eta + \psi)]\}, \quad . \quad \textbf{(37·29)}$$

where $$\sin\psi = \psi \text{ (approximately)} = \frac{3A^2\phi}{4H^2}.$$

On comparing this solution **(37·29)** with the Newtonian one

(37·271) we may characterize the difference which the new theory makes as consisting in a constantly growing change in η and we naturally represent ψ by $-d\eta$. We have for $d\eta$

$$-d\eta = \frac{3A^2\phi}{4H^2}.$$

Now $$\frac{A^2}{H^2} = \frac{2A\kappa M}{h^2} \quad \text{by (37·262)}$$

Therefore $$\frac{A^2}{H^2} = -\frac{4\kappa^2 M^2}{c^2 h^2} \quad \text{by (37·27).}$$

Consequently

$$d\eta = \frac{3\kappa^2 M^2}{c^2 h^2}\phi. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(37·291)}$$

Let us consider what this result means. The angle, η (cf. § 5·5), marks the position of the major axis of the ellipse (or hyperbola or whatever conic it may be) in which, to the first approximation, the planet is travelling. The classical law of Newton makes η a constant; but the new law of Einstein makes it slowly increase; so that we may describe the planetary motion as being along an ellipse (or other conic) whose major axis is moving onwards in the same sense as that of the motion of the planet (cf. § 5·5 in which a very similar problem is treated in great detail).

We may give (37·291) another form, which applies only to elliptical orbits, if we make use of Kepler's law

$$\frac{a^3}{T^2} = \frac{\kappa M}{4\pi^2},$$

a being the semi-major axis of the ellipse and T the period of revolution of the planet (cf. equation 5·54). Now

$$h = \frac{2\pi ab}{T}, \quad (b \equiv \text{semi-minor axis}),$$

and consequently

$$h^2 = \frac{4\pi^2 a^4(1-e^2)}{T^2},$$

so that we obtain finally

$$d\eta = \frac{12\pi^2 a^2}{c^2 T^2(1-e^2)}\phi. \quad . \quad . \quad . \quad . \quad \textbf{(37·292)}$$

§ 37·3. The Advance of Mercury's Perihelion

The formula (37·292) has been confirmed observationally in the case of Mercury, the only planet for which the advance is

appreciable. The relevant data for substitution in the formula are:

$$e = 1/5$$
$$a = 5{\cdot}76 \times 10^{12}\ cm.$$
$$T = 88\ days$$
$$c = 3 \times 10^{10}\ cm./sec.$$

Thus the advance $d\eta$ for a century is, since Mercury makes approximately 415 revolutions in a century,

$$d\eta = \frac{12 \times \pi^2 \times (5{\cdot}76)^2 \times 10^{24} \times 2\pi \times 415}{9 \times 10^{20} \times (88 \times 24 \times 3600)^2 \times (1 - \frac{1}{25})}.$$

This works out to $2{\cdot}046 \times 10^{-4}$ radians or approximately $43''$. The observations show a total advance of $574''$ in a century; but $532''$ can be accounted for by the perturbations due to other planets. There is thus a balance of $42''$ uncounted for by Newton's law and dynamical principles; but admirably explained by Einstein's theory.

§ 37·4. Refraction of Light Passing near the Sun's Disc

We are going to regard light as constituted of something resembling particles (photons) travelling with a velocity which makes ds^2 vanish. Equation **(37·26)** will therefore determine the shapes of the rays of light (tracks of photons) in the gravitational field of the sun; but the constant, H, is now of course infinite so that the equation becomes

$$\frac{d^2u}{d\phi^2} + u = -\frac{3Au^2}{2}. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(37·4)}$$

The term on the right is a very small one and we get the first approximation to a solution of the equation, namely

$$u = Q\cos(\phi - \eta), \quad . \quad . \quad . \quad . \quad \textbf{(37·41)}$$

where Q and η are constants of integration, by ignoring it. This is exactly how we proceeded with **(37·26)**. The greatest value of $u\left(\equiv \frac{1}{r}\right)$ is therefore Q; or the least value of r is $1/Q$ or R. It is convenient to adopt the value zero for η, so that **(37·41)** becomes

$$u = \frac{1}{R}\cos\phi. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(37·411)}$$

The path of a photon according to this approximate solution is a straight line (*abc* in Fig. 37·4 in which R represents the perpendicular distance between the path of the photon and the centre, O, of the sun). To get a better approximation we

naturally substitute **(37·411)** in the small term on the right of **(37·4)**, thus obtaining

$$\frac{d^2u}{d\phi^2} + u = \alpha(1 + \cos 2\phi), \quad . \quad . \quad . \quad \textbf{(37·42)}$$

where

$$\alpha \equiv -3A/4R^2. \quad . \quad . \quad . \quad . \quad \textbf{(37·421)}$$

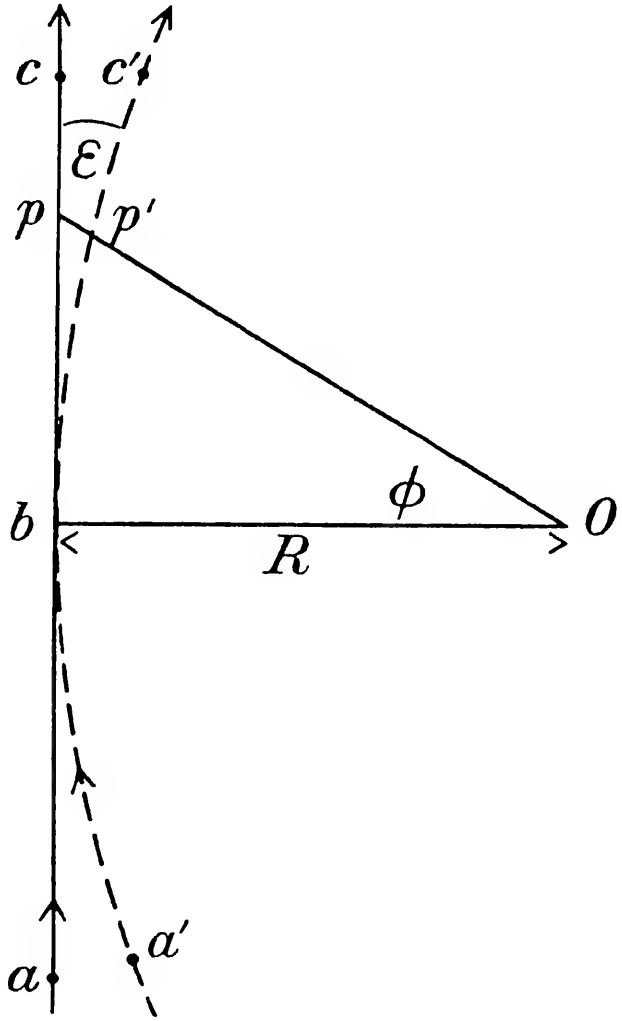

Fig. 37·4

The general solution of this, after assigning the value zero to the constant η, as in **(37·411)** is

$$u = \alpha + \frac{1}{R}\cos\phi - \frac{1}{3}\alpha\cos 2\phi,$$

or

$$u = \frac{4}{3}\alpha + \frac{1}{R}\cos\phi - \frac{2\alpha}{3}\cos^2\phi,$$

and on replacing α by its value given in **(37·421)**

$$u = -\frac{A}{R^2} + \frac{1}{R}\cos\phi + \frac{A}{2R^2}\cos^2\phi,$$

and so, finally, when we substitute the value of A given in **(37·27)** we get

$$u = +\frac{2\kappa M}{c^2R^2} + \frac{1}{R}\cos\phi - \frac{\kappa M}{c^2R^2}\cos^2\phi. \quad . \quad \textbf{(37·43)}$$

We remember that this is the equation of a curve which differs little from a straight line, and that when r is infinite $\cos\phi$ is very little different from zero and $\cos^2\phi$ consequently negligible. Hence when r is infinite ($u = 0$)

$$\cos\phi_\infty = -\frac{2\kappa M}{c^2R},$$

so that the angle, ε (Fig. 37·4), between an asymptote of the real path $a'bc'$ and the straight line abc is

$$\varepsilon = \frac{2\kappa M}{c^2R}$$

approximately and the total deflection of the path of a photon by a gravitating particle of mass M is therefore

$$\zeta = 2\varepsilon$$

or

$$\zeta = \frac{4\kappa M}{c^2R}. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(37·44)}$$

The mass of the sun is approximately 2×10^{33} grams and its radius $6{\cdot}97 \times 10^{10}$ cm. Thus the deflection of the light from a very distant source (star) passing close to the sun's disc is

$$\zeta = \frac{4 \times 6{\cdot}7 \times 10^{-8} \times 2 \times 10^{33}}{9 \times 10^{20} \times 6{\cdot}97 \times 10^{10}}$$

in radians, the constant of gravitation, κ, being taken as

$$6{\cdot}7 \times 10^{-8}\ \text{gm.}^{-1}\ \text{cm.}^3\ \text{sec}^{-2}.$$

Therefore

$$\zeta = 8{\cdot}54 \times 10^{-6} \text{ radians},$$

or

$$\zeta = 4{\cdot}89 \times 10^{-4} \text{ degrees},$$

or finally $\zeta = 1''{\cdot}75$.

This theoretical result was confirmed by the observations carried out by the two British solar eclipse expeditions in 1919. The results obtained by them were:

Sobral expedition $1''{\cdot}98 \pm 0''{\cdot}12$,
Principe expedition $1''{\cdot}61 \pm 0''{\cdot}30$.

The eclipse occurred on May 29th, 1919, a particularly favourable time, since the sun was passing across an area rich in bright stars in the constellation Hyades.

§ 37·5. Displacement of Spectral Lines

The least satisfactory perhaps of the deductions which Einstein made from his gravitational theory was that of a shift of the spectral lines of the light emitted by a distant gravitating body such as a star. The predicted shift is in the direction of longer wave-length and has also been confirmed by observation.

Let dt_s represent the period of some light emitting atomic vibration at the surface of a star of radius r_s and dt_e the corresponding period associated with a similar atom on the earth. Let r_e be the radial distance from the centre of the star to the earth and finally let ds_s and ds_e be the associated intervals. If we regard the atoms as being at rest, we have

$$ds_s^2 = -\left(1 + \frac{A}{r_s}\right)c^2\, dt_s^2,$$

$$ds_e^2 = -\left(1 + \frac{A}{r_e}\right)c^2\, dt_e^2.$$

We now assume—and this is one of the weak places in the deduction—that $ds_s^2 = ds_e^2$. Consequently

$$\left(\frac{dt_s}{dt_e}\right)^2 = 1 + A\left(\frac{1}{r_e} - \frac{1}{r_s}\right),$$

neglecting second order small quantities. Consequently

$$\frac{dt_s}{dt_e} = 1 - \frac{A}{2r_s},$$

since r_e is very great by comparison with r_s.

Now

$$c\, dt_s = \lambda_s,$$
$$c\, dt_e = \lambda_e,$$

where λ_s and λ_e are the corresponding wave-lengths. So that

$$\frac{\lambda_s}{\lambda_e} = 1 + \frac{\kappa M}{c^2 r_s},$$

where M is the mass of the star. Finally therefore

$$\lambda_s - \lambda_e = \frac{\kappa M}{c^2 r_s}\lambda. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(37·5)}$$

If now we assume that the period, which is dt_s at the star's surface, is not modified in the propagation to the earth, the terrestrial observer will find this period in the light which reaches him and he will find its wave-length, λ_s, a little greater than the corresponding wave-length, λ_e, from a similar terrestrial source; the difference being expressed by **(37·5)**.

We can confirm **(37·5)** by a quite different method which invokes the quantum theory. Let us imagine we have two identical photons, one emitted on the star and the other on the earth. Let us call the identical periods, associated with these photons, τ_e. Each then has the energy h/τ_e where h is **Planck's constant** (cf. **§§ 39, 40·3, 40·4**). When the photon from the star has reached the earth its energy will have been reduced by

$$\kappa Mm\left(\frac{1}{r_s} - \frac{1}{r_e}\right)$$

or

$$\kappa Mm/r_s,$$

where M is the mass of the star and m that of the photon. But on reaching the earth its energy must be h/τ_s, where τ_s is the period, as measured by the terrestrial observer, of the light which was emitted on the star. Therefore

$$\frac{h}{\tau_e} - \frac{h}{\tau_s} = \frac{\kappa Mh}{c^2 r_s \tau}$$

since the mass, m, of the photon is its energy divided by c^2, or $h/c^2\tau$. Therefore

$$\tau_s - \tau_e = \frac{\kappa M}{c^2 r_s}\tau,$$

or finally

$$\lambda_s - \lambda_e = \frac{\kappa M\lambda}{c^2 r_s},$$

in agreement with **(37·5)**. It is remarkable that the two calculations lead to the same result since there is a discrepancy between the initial assumptions used with them. The relativistic calculation is based on the assumption of the equality of the *intervals* associated with the vibrating system on the star and on the earth so that the stellar light starts out with its different period. The assumptions of the quantum theory on the other hand assign to the stellar light the same *initial* period as that of the corresponding terrestrial light, the former changing in period as it proceeds on its journey to the earth. The quantum theoretical mode of calculation would seem to be the sounder one.

The earlier attempts to confirm **(37·5)** were uncertain and inconclusive, since the observable quantity is exceedingly small in the case of most stars. More recent observations have been made on very dense stars for which r_s is very small indeed compared with the radii of most stars while the total mass M of a dense star is of the same order of magnitude as that of other stars. These observations leave no doubt about the existence of a

spectral shift agreeing, within the limits of observational error, with what is represented by **(37·5)**.

§ 37·6. Four Identities

We have seen in **§ 35·3** that the conservation of mass, momentum and energy in a medium (or field) is expressed by the vanishing of the divergence of $p_n{}^m$ (or $t_n{}^m$) when there is no impressed force, i.e. no force of external origin. The divergence of this tensor represents, of course, force per unit volume and the principle of equivalence (cf. **§ 36**) requires that the covariant (gravitational) force per unit volume must vanish. So we conclude that the divergence of the appropriate (stress momentum) tensor, $t_n{}^m$, must vanish in a gravitational field. In such a field therefore

$$\frac{\partial t_m{}^p}{\partial x_p} + t_m{}^o\Gamma^p_{op} - t_o{}^p\Gamma^o_{mp} = 0 \quad . \quad . \quad . \quad \textbf{(37·6)}$$

(cf. **§ 36·7**). There are four of these equations, one for each value of m, and since it is the aim of the general theory of relativity to represent mechanical and gravitational phenomena as the outcome of the geometrical properties of the continuum—indeed, it is the aim of relativistic theory to represent all phenomena in this way, but the theory we are describing in this chapter does not go quite so far as that (cf. **§ 36·1**)—we must find some tensor, X^p_m, which we can construct from the fundamental tensors, in particular from the Riemann-Christoffel tensor and whose divergence vanishes identically, and equate it to $t_m{}^p$. In fact in solving **(37·6)** we have to substitute for $t_m{}^p$ a tensor which makes the equation an identity, just as in the solution of a quadratic equation we have to substitute for its x such a number (or numbers) as will reduce it to an identity. This tensor which we are going to identify with $t_m{}^p$ must also have the further property that it agrees with the law **(37·01)** or, possibly and alternatively, **(37·02)** for empty space. The first **tentative** suggestion which occurs to us is

$$G_m{}^p \equiv St_m{}^p, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(37·61)}$$

where S is some constant. This fulfils the condition **(37·01)** in empty space. To be acceptable, however, it should also fulfil the further condition that **div** $G_m{}^p$ should vanish identically. We must therefore investigate **div** $G_m{}^p$. We have

$$\textbf{div}\ G_m{}^p \equiv \frac{\partial G_m{}^p}{\partial x_p} + \Gamma^{\boldsymbol{p}}_{\boldsymbol{po}}G_m{}^o - \Gamma^o_{pm}G_o{}^p,$$

in accordance with **(36·7)** on carrying out the appropriate con-

traction. In order to simplify and shorten the investigation we shall introduce co-ordinates which are orthogonal in a small or restricted neighbourhood, so that in such a neighbourhood (cf. § 36·9) the first derivatives of the $g_{\mu\nu}$ vanish and consequently

$$\textbf{div}\ G_m{}^p = \frac{\partial G_m{}^p}{\partial x_p}. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(37·62)}$$

Now (cf. **37·01**)

$$G_m{}^p \equiv g^{pn}\frac{\partial \Gamma^r_{rm}}{\partial x_n} - g^{pn}\frac{\partial \Gamma^r_{mn}}{\partial x_r} - g^{pn}\Gamma^r_{ro}\Gamma^o_{mn} + g^{pn}\Gamma^r_{no}\Gamma^o_{mr}. \quad . \quad \textbf{(37·621)}$$

Therefore in the small neighbourhood in question

$$G_m{}^p \equiv \frac{1}{2}g^{pn}g^{rt}\frac{\partial^2 g_{rt}}{\partial x_n\,\partial x_m} - \frac{1}{2}g^{pn}g^{rt}\frac{\partial^2 g_{tm}}{\partial x_r\,\partial x_n} - \frac{1}{2}g^{pn}g^{tr}\frac{\partial^2 g_{tn}}{\partial x_r\,\partial x_m} + \tfrac{1}{2}g^{pn}g^{tr}\frac{\partial^2 g_{mn}}{\partial x_r\,\partial x_t}, \quad . \quad . \quad \textbf{(37·63)}$$

since the first derivatives of the g_{mn} vanish. We must not, of course, in calculating **div** $G_m{}^p$, depend on this approximation for $G_m{}^p$ until we have assured ourselves that the vanishing terms in **(37·621)** contribute nothing to the divergence. It is easy to see that this is the case and thus

$$\textbf{div}\,G_m{}^p \equiv \frac{1}{2}g^{pn}g^{rt}\frac{\partial^3 g_{rt}}{\partial x_p\,\partial x_n\,\partial x_m} - \frac{1}{2}g^{pn}g^{rt}\frac{\partial^3 g_{tm}}{\partial x_p\,\partial x_r\,\partial x_n} - \frac{1}{2}g^{pn}g^{rt}\frac{\partial^3 g_{tn}}{\partial x_p\,\partial x_r\,\partial x_m} + \frac{1}{2}g^{pn}g^{rt}\frac{\partial^3 g_{mn}}{\partial x_p\,\partial x_r\,\partial x_t}.$$

It is easily seen that two of the terms in this expression annul one another, so that

$$\textbf{div}\,G_m{}^p \equiv \frac{1}{2}g^{pn}g^{rt}\frac{\partial^3 g_{rt}}{\partial x_p\,\partial x_n\,\partial x_m} - \frac{1}{2}g^{pn}g^{rt}\frac{\partial^3 g_{tn}}{\partial x_p\,\partial x_r\,\partial x_m}, \quad . \quad \textbf{(37·64)}$$

The divergence of G_m^p therefore does not vanish identically and the suggestion **(37·61)** fails in consequence.

Let us next investigate the divergence of $g^m_p G$, where the scalar, G, is defined by

$$G = g^{ab}G_{ab}.^1$$

In a restricted neighbourhood or region of space-time

$$G = \frac{1}{2}g^{ab}g^{rt}\frac{\partial^2 g_{rt}}{\partial x_b\,\partial x_a} - \frac{1}{2}g^{ab}g^{rt}\frac{\partial^2 g_{at}}{\partial x_r\,\partial x_b} - \frac{1}{2}g^{ab}g^{rt}\frac{\partial^2 g_{bt}}{\partial x_r\,\partial x_a} + \frac{1}{2}g^{ab}g^{rt}\frac{\partial^2 g_{ab}}{\partial x_r\,\partial x_t}. \quad . \quad . \quad \textbf{(37·65)}$$

[1] Such a scalar as $T = g^{ab}T_{ab}$ is called a **Laue's scalar** (cf. § 35·2.)

Hence

$$\mathbf{div}\, g_m{}^p G = \frac{1}{2} g^{ab} g^{rt} \frac{\partial^3 g_{rt}}{\partial x_m\, \partial x_b\, \partial x_a} - \frac{1}{2} g^{ab} g^{rt} \frac{\partial^3 g_{at}}{\partial x_m\, \partial x_r\, \partial x_b} - \frac{1}{2} g^{ab} g^{rt} \frac{\partial^3 g_{bt}}{\partial x_m\, \partial x_r\, \partial x_a} + \frac{1}{2} g^{ab} g^{rt} \frac{\partial^3 g_{ab}}{\partial x_m\, \partial x_r\, \partial x_t}.$$

When we remember that we may replace any index which represents a summation, for example a, by any other letter not already employed in some other sense we may write the last result in the form:

$$\mathbf{div}\, g_m{}^p G = g^{ab} g^{rt} \frac{\partial^3 g_{rt}}{\partial x_m\, \partial x_a\, \partial x_b} - g^{ab} g^{rt} \frac{\partial^3 g_{at}}{\partial x_m\, \partial x_r\, \partial x_b},$$

or finally

$$\mathbf{div}\, g_m{}^p G = g^{pn} g^{rt} \frac{\partial^3 g_{rt}}{\partial x_p\, \partial x_n\, \partial x_m} - g^{pn} g^{rt} \frac{\partial^3 g_{tn}}{\partial x_p\, \partial x_r\, \partial x_m}. \quad . \quad \textbf{(37·66)}$$

On comparing this with **(37·64)** we see that

$$\mathbf{div}\, (G_m{}^p - \tfrac{1}{2} g_m{}^p G) = 0 \quad . \quad . \quad . \quad \textbf{(37·67)}$$

identically. We have deduced it by using co-ordinates for which the g_{mn} are constant in a limited region of space-time. Since it is, however, a tensor relationship it must be universally valid. Furthermore it is evident that

$$\mathbf{div}\, \{G_m{}^p - \tfrac{1}{2} g_m{}^p (G + C)\} = 0, \quad . \quad \textbf{(37·671)}$$

where C is any constant. Each of the statements **(37·67)** and **(37·671)** represents four identities, one for each value of m.

We now realize that the tentative suggestion, **(37·61)**, must be abandoned in favour of one or other of the following suggestions:

$$G_m{}^p - \tfrac{1}{2} g_m{}^p G = S t_m{}^p, \quad . \quad . \quad \textbf{(37·68)}$$

$$G_m{}^p - \tfrac{1}{2} g_m{}^p (G + C) = S t_m{}^p, \quad . \quad . \quad \textbf{(37·681)}$$

where S and C are constants which remain to be determined.

§ 37·7. The General Law of Gravitation

The results obtained in the preceding section suggest, as we have seen, the two alternatives **(37·68)** and **(37·681)** for the general law of gravitation. The former is in agreement with Einstein's law **(37·01)** for empty space, namely

$$G_m{}^p = 0,$$

or

$$G_{mn} = 0.$$

This can be seen by multiplying **(37·68)** by g^m_p and carrying out the summation. Thus:

$$g_p{}^m G_m{}^p - \tfrac{1}{2} g_p{}^m g_m{}^p G = S g_p{}^m t_m{}^p,$$

and hence

$$G - 2G = St,$$

where $$t \equiv g_p{}^m t_m{}^p \equiv g^{mn} t_{mn}.$$

Therefore $$-G = St,$$

so that G vanishes in empty space and consequently so do G_m^p and G_{mn}.

If we prefer for empty space the law **(37·02)**, i.e.

$$G_{mn} = \lambda g_{mn},$$

λ being a small constant usually called the **cosmic constant,** we adopt the second alternative, namely **(37·681).** On multiplying both sides of this equation by $g_p{}^m$ and summing we get

$$G - 2(G + C) = St,$$

and in empty space therefore

$$G = -2C. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(37·7)}$$

Hence in empty space

$$G_m{}^p - \tfrac{1}{2} g_m{}^p (-C) = 0,$$

or $$G_m{}^p = -\tfrac{1}{2} C g_m{}^p,$$

and $$G_{mn} = -\tfrac{1}{2} C g_{mn}.$$

This will coincide with **(37·02)** when we assign to C the value

$$C = -2\lambda, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(37·71)}$$

so that in empty space

$$G = 4\lambda, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(37·72)}$$

(cf. **37·7**) and **(37·681)** may be written

$$G_m{}^p - \tfrac{1}{2} g_m{}^p (G - 2\lambda) = St_m{}^p. \quad . \quad . \quad \textbf{(37·73)}$$

The significance of the constant S and the value to be assigned to it will be investigated in **§ 37·9.**

§ 37·8. Isotropic Co-ordinates

We shall now search for such co-ordinates, for the field of an isolated particle, as give the square of the interval the form

$$ds^2 = \phi_1(\rho)\{d\rho^2 + \rho^2\, d\theta^2 + \rho^2 \sin^2\theta\, d\phi^2\} - \phi_2(\rho) c^2\, dt^2, \qquad \textbf{(37·8)}$$

instead of

$$ds^2 = \left(1 - \frac{A}{r}\right) dr^2 + r^2\, d\theta^2 + r^2 \sin^2\theta\, d\phi^2 - \left(1 + \frac{A}{r}\right) c^2\, dt^2, \qquad \textbf{(37·81)}$$

where A means the constant $-2\kappa M/c^2$, of **(37·27).** We shall see as we proceed that this is possible.

It will be observed that in the latter expression we have written

$$1 - A/r \text{ for } 1/(1 + A/r).$$

This is permissible when the square and higher powers of A/r may be neglected.

We have from **(37·8)** and **(37·81)**

$$\phi_1(\rho)d\rho^2 = \left(1 - \frac{A}{r}\right)dr^2$$

and

$$\phi_1(\rho)\rho^2 = r^2,$$

hence

$$\frac{d\rho}{\rho} = \left(1 - \frac{A}{2r}\right)\frac{dr}{r} = \frac{dr}{r} - \frac{A}{2}\frac{dr}{r^2},$$

when the square and higher powers of A/r may be neglected. Consequently to this degree of approximation

$$d\log\rho = d\log r + d\,(A/2r)$$

and

$$\log\rho = \log r + A/2r + C,$$

where C is a constant.

Therefore

$$\frac{\rho}{r} = De^{A/2r} = D(1 + A/2r)$$

approximately, D being the constant e^C. If we fix this to be unity we get

$$r = \rho - \frac{A}{2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(37·82)}$$

Consequently we may write, whenever the square and higher powers of A/r may be ignored,

$$ds^2 = \left(1 - \frac{A}{\rho}\right)\{dx^2 + dy^2 + dz^2\} - \left(1 + \frac{A}{\rho}\right)c^2\,dt^2, \quad . \quad \textbf{(37·83)}$$

in which $d\rho^2 + \rho^2\,d\theta^2 + \rho^2\sin^2\theta\,d\phi^2$ has been replaced by $dx^2 + dy^2 + dz^2$.

It will be noticed that the velocity of light, in these co-ordinates, has the same value, namely $(1 + A/\rho)c$ or $(1 - 2\kappa M/c^2\rho)c$ in all directions.

We shall call these new co-ordinates **isotropic co-ordinates** although strictly speaking they only represent what isotropic co-ordinates reduce to when A^2/ρ^2 and higher powers of A/ρ may be ignored. They will be used in the problem of determining the constant S of **§ 37·7.**

§ 37·9. Complete Statement of the Law of Gravitation

To complete the statement of the law of gravitation we have to evaluate the constant, S, of **§ 37·7.** To achieve this we

employ the device of passing to the limit where space-time becomes Galilean, or practically Galilean, and it is a help to replace $-c^2\,dt^2$ by dw^2 in the expression for ds^2 in **(37·83)**. We shall also temporarily replace dx_1, dx_2, and dx_3 by dx, dy, and dz, as we did in **(37·83)**. This expression then becomes

$$ds^2 = \left(1 - \frac{A}{r}\right)(dx^2 + dy^2 + dz^2) + \left(1 + \frac{A}{r}\right)dw^2,$$

since we are concerned with the state of affairs at a very distant point where ρ approaches r and where the geometry is approaching the Galilean limit for which

$$r^2 = x^2 + y^2 + z^2.$$

On making use of **(37·63)** we obtain

$$G_1{}^1 = \frac{\partial^2}{\partial x^2}\left(\frac{g_{22} + g_{33}}{2} + \tfrac{1}{2}g_{44}\right) + \frac{\partial^2}{\partial y^2}(\tfrac{1}{2}g_{11}) + \frac{\partial^2}{\partial z^2}(\tfrac{1}{2}g_{11}) \quad . \quad \textbf{(37·9)}$$

and the symmetry which the isotropic co-ordinates manifest has the consequence that

$$G_1{}^1 = G_2{}^2 = G_3{}^3$$

and

$$g_{11} = g_{22} = g_{33} = 1 - A/r.$$

Similarly we find

$$G_4{}^4 = \tfrac{1}{2}\nabla^2 g_{44}. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(37·91)}$$

Turning back to **(37·9)** let us substitute in it for g_{11}, g_{22}, g_{33}, and g_{44} their values as given in the expression for ds^2. We thus obtain

$$G_1{}^1 = \frac{\partial^2}{\partial x^2}\left(\frac{3}{2} - \frac{A}{2r}\right) + \frac{\partial^2}{\partial y^2}\left(\frac{1}{2} - \frac{A}{2r}\right) + \frac{\partial^2}{\partial z^2}\left(\frac{1}{2} - \frac{A}{2r}\right).$$

We may just as well write this in the form

$$\left.\begin{aligned} G_1{}^1 = G_2{}^2 = G_3{}^3 = -\tfrac{1}{2}\nabla^2(A/r) = \nabla^2\left(\frac{\kappa M}{c^2 r}\right), \\ \text{while} \\ G_4{}^4 = +\tfrac{1}{2}\nabla^2(A/r) = \nabla^2\left(-\frac{\kappa M}{c^2 r}\right). \end{aligned}\right\} \quad . \quad \textbf{(37·92)}$$

These expressions apply to a point in empty space at a great distance, r, from a mass M and Einstein's law of gravitation for points in empty space is $G_m{}^p = 0$, i.e.

$$G_1{}^1 = 0,\ G_2{}^2 = 0, \text{ etc.} \quad . \quad . \quad . \quad \textbf{(37·921)}$$

It follows therefore *from Einstein's* law that

$$\nabla^2\left(\frac{\kappa M}{r}\right) = 0.$$

Thus Einstein's law agrees with Laplace's formula when r is very great. We may note here that **(37·92)** is equivalent to

$$\left.\begin{aligned} G_1{}^1 &= +\tfrac{1}{2}\nabla^2 g_{11}, \\ G_2{}^2 &= +\tfrac{1}{2}\nabla^2 g_{22}, \\ G_3{}^3 &= +\tfrac{1}{2}\nabla^2 g_{33}, \\ G_4{}^4 &= +\tfrac{1}{2}\nabla^2 g_{44}, \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad \textbf{(37·922)}$$

since the additive part in each g contributes nothing to the Laplacian. The variable part of g_{11}, g_{22}, or g_{33} is $2\kappa M/c^2 r$. Since the partial differential equations **(37·921)** are linear we obviously have for a distant point in the field of several masses M_1, M_2, etc.

$$\left.\begin{aligned} &G_1{}^1 = G_2{}^2 = G_3{}^3 = \nabla^2\left(\Sigma\frac{\kappa M}{c^2 r}\right) = -\frac{1}{c^2}\nabla^2 V = 0 \\ &\text{and} \qquad G_4{}^4 = \nabla^2\left(-\Sigma\frac{\kappa M}{c^2 r}\right) = +\frac{1}{c^2}\nabla^2 V = 0. \end{aligned}\right\} \quad \textbf{(37·93)}$$

Thus the variable part of g_{11}, g_{22}, or g_{33} is $-2V/c^2$ and that of g_{44} is $+2V/c^2$. Therefore

$$\left.\begin{aligned} g_{11} = g_{22} = g_{33} &= 1 - 2V/c^2, \\ \text{and} \qquad g_{44} &= 1 + 2V/c^2. \end{aligned}\right\} \quad . \quad . \quad \textbf{(37·931)}$$

The function $V\left(= -\kappa\Sigma\frac{M}{r}\right)$ is the Newtonian gravitational potential.

We are now in a position to consider the case where a continuously distributed mass exists at and in the neighbourhood of the point at which the functions $G_1{}^1$, etc., are being evaluated. The partial differential equations for Einstein's law of gravitation are now (cf. **37·68**)

$$G_m{}^p - \tfrac{1}{2}g_m{}^p G = St_m{}^p$$

and therefore

$$\left.\begin{aligned} G_1{}^1 - \tfrac{1}{2}g_1{}^1 G &= St_1{}^1, \\ G_4{}^4 - \tfrac{1}{2}g_4{}^4 G &= St_4{}^4. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad \textbf{(37·94)}$$

It is clear that in the limiting Newtonian case each of these must pass into Poisson's equation

$$\nabla^2 V = 4\pi\kappa\rho,$$

where ρ is the mass density. A word should be said here about the **t**. It is in fact the tensor whose components appear in equations **(10·55)** (with the opposite sign). It is a tensor which is associated with gravitational forces only. All its components —in the special limiting case we are studying—except $t_4{}^4$ vanish. This is because the limiting case we are dealing with is one in which all the masses involved are at rest and in which there are

no time variations. Clearly $t_4{}^4$ will be identical with the energy per unit volume

$$t_4{}^4 = \rho c^2 \quad . \quad . \quad . \quad . \quad . \quad (37{\cdot}95)$$

where ρ is the mass density. Now consider the equations **(37·94)**. By **(37·93)**

$$G_1{}^1 = -\frac{4\pi\kappa\rho}{c^2}$$

and

$$G_4{}^4 = +\frac{4\pi\kappa\rho}{c^2}$$

and $G = g_m{}^p G_p{}^m$ and therefore in our special case

$$G = G_1{}^1 + G_2{}^2 + G_3{}^3 + G_4{}^4$$

or

$$G = -\frac{8\pi\kappa\rho}{c^2}.$$

Substituting in the first of the equations **(37·94)** we have

$$-\frac{4\pi\kappa\rho}{c^2} - \frac{1}{2}\left(-\frac{8\pi\kappa\rho}{c^2}\right) = St_1{}^1.$$

Hence as we have anticipated $t_1{}^1$ vanishes. On substituting in the last of equations **(37·94)** we get

$$+\frac{4\pi\kappa\rho}{c^2} - \frac{1}{2}\left(-\frac{8\pi\kappa\rho}{c^2}\right) = St_4{}^4$$

or

$$\frac{8\pi\kappa\rho}{c^2} = S\rho c^2$$

and therefore

$$\boxed{S = \frac{8\pi\kappa}{c^4}.} \quad . \quad . \quad . \quad . \quad . \quad (37{\cdot}96)$$

Thus Einstein's general law of gravitation becomes

$$G_n{}^m - \tfrac{1}{2}g_n{}^m G = \frac{8\pi\kappa}{c^4} t_n{}^m. \quad . \quad . \quad . \quad (37{\cdot}97)$$

The alternative law is

$$G_n{}^m - \tfrac{1}{2}g_n{}^m(G - 2\lambda) = \frac{8\pi\kappa}{c^4} t_n{}^m. \quad . \quad . \quad (37{\cdot}98)$$

§ 38. Electromagnetic Field Equations

In the limiting case of Galilean space and orthogonal co-ordinates one set of the four field equations of electromagnetism must approach the form **(34·01)** or **(34·012)**. The general form of this set of equations must therefore be

$$\frac{1}{g^{1/2}} \frac{\partial(g^{1/2} F^{\alpha\beta})}{\partial x_\beta} = s^\alpha. \quad . \quad . \quad . \quad . \quad (38)$$

The remaining set of four equations **(34·02)** will not change its form. The reason for this is very simple. A typical equation of the set **(34·02)** is

$$\frac{\partial F_{\alpha\beta}}{\partial x_\gamma} + \frac{\partial F_{\gamma\alpha}}{\partial x_\beta} + \frac{\partial F_{\beta\gamma}}{\partial x_\alpha} = 0 \quad . \quad . \quad . \quad \textbf{(38·01)}$$

and as $F_{\alpha\beta}$ can be put in the form $\partial A_\beta/\partial x_\alpha - \partial A_\alpha/\partial x_\beta$ the left-hand member of **(38·01)** is a tensor not merely within the scope of special relativity, but quite generally (cf. **§ 36·6**).

In order to get the relationship between $F_{\alpha\beta}$ and $F^{\alpha\beta}$ we have merely to identify the tensors $g_\beta{}^\alpha$, $g_{\alpha\beta}$, and $g^{\alpha\beta}$ in **§ 35·6** with the corresponding fundamental tensors of the general theory of relativity. With this identification **(35·63)** will hold universally in any co-ordinate system.

§ 38·1. The Electromagnetic Stress-Momentum-Energy Tensor

We may introduce at once a tensor

$$t_\alpha{}^\mu \equiv F_{\alpha\beta}F^{\beta\mu} + \tfrac{1}{4}g_\alpha{}^\mu F^{\sigma\tau}F_{\sigma\tau}, \quad . \quad . \quad . \quad . \quad \textbf{(38·1)}$$

corresponding to **(35·26)** with the $\delta_\alpha{}^\mu$ naturally replaced by $g_\alpha{}^\mu$. Indeed, if we start out with the equation corresponding to **(35·222)**, namely

$$f_\alpha = \frac{1}{g^{1/2}} F_{\alpha\beta}\frac{\partial(g^{1/2}F^{\beta\mu})}{\partial x_\mu}, \quad . \quad . \quad . \quad . \quad \textbf{(38·11)}$$

we shall in fact find

$$f_\alpha = \frac{\partial t_\alpha{}^\mu}{\partial x_\mu} + \Gamma^\mu_{\omega\mu}t_\alpha{}^\omega - \Gamma^\omega_{\alpha\mu}t_\omega{}^\mu, \quad . \quad . \quad . \quad \textbf{(38·12)}$$

which is the appropriate generalization of **(35·251)**. We may establish it simply by noting that it is true in a restricted region in which co-ordinates can be so chosen that the g_{mn} are constants and being a tensor relationship it must be independent of the co-ordinate system.

§ 38·2. Conservation of Electric Charge

The antisymmetrical character of the tensor $F^{\alpha\beta}$ in **(38)** makes it evident that

$$\frac{1}{g^{1/2}}\frac{\partial}{\partial x_\alpha}(g^{1/2}s^\alpha) = 0. \quad . \quad . \quad . \quad . \quad \textbf{(38·2)}$$

This is the generalization of **(22·75)** which expresses the conservation of electricity.

§ 38·3. The Problem of the Unification of Gravitational and Electrical Phenomena

The general theory of relativity which we have been describing represents gravitational phenomena as manifestations of the geometrical properties of the space-time continuum. Other phenomena, those of electromagnetism in particular, merely occur within the Riemannian space-time framework and are not associated with it in any more intimate way. The theory which accounts so well, not only for the character, but for the very existence of gravitational phenomena, gives no reason for the existence of electrical phenomena. In **§ 36·1** we have given a brief outline of one of the suggestions—that of Kaluza—for widening Einstein's gravitational theory in such a way that not only gravitational phenomena, but also electrical phenomena, emerge as expressions of the metrical features of the continuum.

We shall now give an outline of another suggestion which is due to H. Weyl. In studying parallel displacement **(§ 36·5)** we made the assumption that the absolute value of the displaced interval, **ds**, or its square $\mathbf{ds}^2$, was not altered by the displacement. Later, in **§§ 36·8** and **36·9**, we found that the components of the displacement vector, **ds**, were in general changed when the vector was taken, by parallel displacement, round a closed loop—the special case of no change in the values of the individual components, i.e. the special case in which the direction of **ds** was unaffected as well as its absolute value, occurring when a certain tensor, the Riemann-Christoffel tensor, vanished. Now Weyl's suggestion is that not only may the *components* of **ds** (or ξ) suffer a change when it is taken by parallel displacement round a closed loop, *but also its absolute value*. Leaving on one side for the moment the difficulty of attaching a clear meaning to this suggestion, we may note first it amounts to saying that $\oint \delta\mathbf{ds}$ (or $\oint \delta\xi$) in general differs from zero or, what amounts to the same thing, $\int_1^2 \delta\xi$ depends on the path of the parallel displacements of ξ from 1 to 2. The simplest form that $\delta\xi$ can take is

$$\delta\xi = \xi\phi_\alpha \delta x_\alpha, \qquad . \quad . \quad . \quad . \quad . \quad \textbf{(38·3)}$$

in which ϕ_α is a new (covariant) vector and summation with respect to α is implied.

Therefore

$$\oint \delta \log \xi = \oint \phi_\alpha \delta x_\alpha$$

$$= \frac{1}{2}\iint \left\{ \frac{\partial \phi_\nu}{\partial x_\mu} - \frac{\partial \phi_\mu}{\partial x_\nu} \right\} dS^{\mu\nu} \quad . \quad . \quad \textbf{(38·31)}$$

by Stokes' theorem. Weyl identifies this new vector ϕ with the electromagnetic vector potential **(§ 35·6)** and his theory thus furnishes or appears to furnish a reason for the existence of electromagnetic phenomena.

Weyl's assumption bristles with difficulties. Let us first interpret in the sense in which it was apparently intended to be interpreted when it was originally made. We think of ξ as a displacement in space-time determined by a rigid rod and a clock. On subjecting it to a parallel displacement its absolute value is changed by hypothesis and we may not, indeed we cannot, use this physical thing—rod cum clock—to determine it on account of the ambiguity of $\int_1^2 \delta\xi$. We are forced therefore to have a physical device—a **gauge**—at every place in space-time to decide what the values of intervals, **ds**, are. Thus Weyl's theory, interpreted in this way, renders an account of the vector ϕ_α, i.e. it renders an account of electromagnetic phenomena in terms of a purely imaginary phenomenon, or at any rate in terms of a phenomenon of the existence of which there is not the faintest observational evidence. Weyl's theory, therefore, always provided the interpretation of it just given is the right one, fails to perform its chief function, which is to connect electromagnetic phenomena with other physical phenomena, more especially those of gravitation.

There is perhaps more to be said in favour of Weyl's theory when we regard his space-time as a conceptual one in which physical phenomena are represented, as Eddington would say, graphically—when we regard it as having the same relationship to phenomena as a plane surface, on which p is plotted against v, has to Boyle's law. Eddington adopts this view and has gone some way beyond Weyl. His conceptual space-time is built up on the parallel displacement formula

$$\delta\xi^a = -\gamma^a_{bc}\xi^b \delta x_c, \quad . \quad . \quad . \quad . \quad \textbf{(38·32)}$$

in which γ^a_{bc} generally differs from our Γ^a_{bc}, and

$$ds^2 = g_{mn}\xi^m\xi^n. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(38·33)}$$

From these we can construct (as in **§ 36·8)** a widened Riemann-Christoffel tensor

$$*B^r_{mn,\,k} \equiv \frac{\partial \gamma^r_{km}}{\partial x_n} - \frac{\partial \gamma^r_{mn}}{\partial x_k} - \gamma^r_{ko}\gamma^o_{mn} + \gamma^r_{ro}\gamma^o_{kn}. \quad . \quad \textbf{(38·35)}$$

Just as in the Riemannian continuum we have a tensor, $*G_{mn}$, which we get by contraction

$$*G_{mn} \equiv *B^r_{mn,\,r} \equiv \frac{\partial \gamma^r_{rm}}{\partial x_n} - \frac{\partial \gamma^r_{mn}}{\partial x_r} - \gamma^r_{ro}\gamma^o_{mn} + \gamma^r_{no}\gamma^o_{mr}. \quad . \quad \textbf{(38·36)}$$

There is however another mode of contracting $*B^r_{mn,\,k}$ which, unlike the corresponding Riemannian contraction **(37·011),** does not in general yield a mere identity, namely

$$*F_{nk} \equiv *B^r_{rnk} \equiv \frac{\partial \gamma^r_{kr}}{\partial x_n} - \frac{\partial \gamma^r_{nr}}{\partial x_k}. \quad . \quad . \quad . \quad \textbf{(38·37)}$$

Eddington identifies $*G_{mn}$ with λg_{mn} where λ is a constant. Thus he gets a law of gravitation. The new tensor F_{nk} he naturally identifies with the covariant electromagnetic field vector. A full account of Eddington's theory will be found in his Mathematical Theory of Relativity. Notwithstanding its attractiveness, the opinion may be expressed here that Kaluza's theory is the correct one, though this does not command universal assent. There seems to be some scope for the use of Weyl's theory in a quite different sphere. According to H. T. Flint it promises, when associated with a certain imaginary gauge system, to furnish a basis for wave-mechanics.

CHAPTER VII

COSMOLOGICAL SPECULATIONS

§ 38·4. Einstein's Cylindrical World

THE treatment of planetary motion given in §§ **37·1** and **37·2** by Einstein's law of gravitation led to Galilean values for the g_{mn} at infinity and to an infinite spacial world. In such a world energy will be constantly passing out to infinity in the form of radiation and (in a world containing a finite total mass at any rate) never returning. And, indeed, it is difficult to avoid the conclusion that masses, even whole stars, will travel away into infinity never to return. Reflections like these worried Einstein and in consequence he devised a world which was spacially finite. He imagined the spacial part of the world only (not space-time) to be spherical (in a sense which will be explained later) and that matter is uniformly spread through it. The latter supposition is to be understood in the same sense as that in which a gas may be said to spread uniformly through the volume of a containing vessel. We understand by it that equal volumes in the vessel contain equal masses of the gas, but only so long as these equal volumes are sufficiently large to contain large numbers of molecules.

The term ' spherical ' is used to indicate that the mathematical description of the world is like that of the 2-dimensional continuum which constitutes the surface of a sphere in Euclidean space. The points of any 2-dimensional continuum can be specified by two independent co-ordinates. But we are at liberty to use three or more co-ordinates for a 2-dimensional continuum. There will of course be one or more relationships between them when their number exceeds two. If in specifying the points of such a continuum we use three co-ordinates, x_1, x_2, x_3, which are subject to the condition

$$R^2 = x_1^2 + x_2^2 + x_3^2,$$

where R is a constant, then the mathematical properties of the continuum are identical with those of the surface of a sphere in Euclidean space. Now Einstein's world (i.e. the spacial part of

it, or the world at an instant) can be described by four co-ordinates, x_1, x_2, x_3, x_4, subject to the condition

$$R^2 = x_1^2 + x_2^2 + x_3^2 + x_4^2, \quad . \quad . \quad . \quad \textbf{(38·4)}$$

where R is a constant. It must be carefully borne in mind that we are now concerned with space only, a 3-dimensional continuum, and not for the moment with the 4-dimensional continuum of space and time. So that the co-ordinate, x_4, in **(38·4)** has nothing to do with the time. Since, however, we have been using x_1, x_2, x_3, and x_4 as the independent co-ordinates of points in the space-time continuum and shall continue this practice, it will be well to replace x_4 in **(38·4)** by another symbol, ξ, which we shall eventually eliminate, leaving x_1, x_2, and x_3 as independent spacial co-ordinates. We shall write therefore

$$R^2 = x_1^2 + x_2^2 + x_3^2 + \xi^2. \quad . \quad . \quad . \quad \textbf{(38·41)}$$

Further

$$d\sigma^2 = dx_1^2 + dx_2^2 + dx_3^2 + d\xi^2, \quad . \quad . \quad \textbf{(38·42)}$$

where $d\sigma$ is of course the spacial interval and should not be confused with ds. To eliminate ξ let us write

$$\xi = \sqrt{R^2 - x_1^2 - x_2^2 - x_3^2}.$$

Therefore

$$d\xi = \frac{-x_1\,dx_1 - x_2\,dx_2 - x_3\,dx_3}{\sqrt{R^2 - x_1^2 - x_2^2 - x_3^2}},$$

and thus

$$d\sigma^2 = dx_1^2 + dx_2^2 + dx_3^2 + \frac{(x_1\,dx_1 + x_2\,dx_2 + x_3\,dx_3)^2}{R^2 - r^2}, \quad \textbf{(36·43)}$$

where $$r^2 \equiv x_1^2 + x_2^2 + x_3^2,$$

and therefore

$$d\sigma^2 = \left\{\delta_{mn} + \frac{x_m x_n}{R^2 - r^2}\right\} dx_m\,dx_n, \quad . \quad . \quad \textbf{(38·44)}$$

where $$\delta_{mn} = \begin{cases} 1, & m = n, \\ 0, & m \neq n, \end{cases}$$

and the summation convention is in operation.

Finally, then, when we turn to the space-time continuum we have

$$ds^2 = g_{mn}\,dx_m\,dx_n,$$

and

$$\left.\begin{aligned} g_{mn} &= \delta_{mn} + \frac{x_m x_n}{R^2 - r^2} \\ &\text{when } m \text{ and } n \text{ range over the values 1, 2, and 3 ; while} \\ g_{4m} &= 0, \text{ for } m = 1,\ 2, \text{ and } 3 \\ &\text{and} \\ g_{44} &= 1. \end{aligned}\right\} \quad \textbf{(38·45)}$$

The value unity has been assigned to g_{44} because it seems to be simpler to replace $-c^2\,dt^2$ by dw^2.

Now of course we may not adopt values for the g_{mn} just according to our caprice, as we appear to have done. They must satisfy a law of gravitation which is either identical with one or other of the two statements **(37·68)** or **(37·681)** or differs sufficiently little from them. Einstein showed, in a way we shall describe, that the coefficients g_{mn}, as given by **(38·45)**, satisfy the latter of the two laws of gravitation provided that we may regard space as filled with matter uniformly distributed and at rest. For anything we know at present it is permissible to adopt this view of the world so long as we deal only with sufficiently large volumes, just in the same way as it is possible to regard the density of a gas as having the same value in all parts of the containing vessel. There is also no difficulty, when we treat the world in this large-scale fashion, in regarding the matter filling it as at rest. After all, nobody would be able to notice in the course of a lifetime any departure of the stellar framework from perfect rigidity unless he were provided with comparatively elaborate means of observation. He would only notice the quite unimportant planetary movements.

The law

$$G_m{}^p - \tfrac{1}{2}g_m{}^p(G - 2\lambda) = \frac{8\pi\kappa}{c^4}t_m{}^p, \quad . \quad . \quad . \quad \textbf{(38·46)}$$

we shall show, is satisfied for such a world **provided a suitable value, depending on the uniform density, ρ, be assigned to the constant λ.**

Now by **(38·45)** the g_{mn} approach the values

$$g_{11} = g_{22} = g_{33} = g_{44} = 1$$

at all points where $x_1 = x_2 = x_3 = 0$, the remaining components vanishing at such points. And of course the contravariant components, g^{11}, g^{22}, g^{33}, and g^{44} are likewise equal to unity, the remaining ones vanishing. Further, the differential quotients $\partial g_{mn}/\partial x_r$ vanish in a small neighbourhood containing the point $x_1 = x_2 = x_3 = 0$. In fact, in the neighbourhood in question

$$\left.\begin{aligned} \frac{\partial g_{mn}}{\partial x_m} &= \frac{x_n}{R^2 - r^2}, \quad m \neq n \\ &\text{and } m, n, = 1, 2, 3, \\ \frac{\partial g_{mm}}{\partial x_m} &= \frac{2x_m}{R^2 - r^2}, \quad m = 1, 2, 3, \end{aligned}\right\} \quad . \quad \textbf{(38·47)}$$

the remaining differential quotients vanishing absolutely. The second differential quotients are

$$\left.\begin{aligned}\frac{\partial^2 g_{mn}}{\partial x_m\,\partial x_n} &= \frac{1}{R^2 - r^2} = \frac{1}{R^2}\\ \frac{\partial^2 g_{mm}}{\partial x_m{}^2} &= \frac{2}{R^2 - r^2} = \frac{2}{R^2}\end{aligned}\right\}\begin{matrix}\text{at the point } r = 0\\ \text{when } m,\ n = 1,\\ 2,\ 3 \text{ and no sum-}\\ \text{mation.}\end{matrix} \quad (38{\cdot}471)$$

All other second derivatives of the g_{mn} vanish.

We now substitute in **(37·63)**, i.e. in

$$G_m{}^p \equiv \tfrac{1}{2}g^{pn}g^{rs}\frac{\partial^2 g_{rs}}{\partial x_m\,\partial x_n} - \tfrac{1}{2}g^{pn}g^{rs}\frac{\partial^2 g_{sm}}{\partial x_r\,\partial x_n} - \tfrac{1}{2}g^{pn}g^{rs}\frac{\partial^2 g_{sn}}{\partial x_m\,\partial x_r} + \tfrac{1}{2}g^{pn}g^{rs}\frac{\partial^2 g_{mn}}{\partial x_r\,\partial x_s} \quad . \quad . \quad (38{\cdot}472)$$

and easily find

$$\left.\begin{aligned}G_1{}^1 = G_2{}^2 = G_3{}^3 &= -\frac{2}{R^2},\\ G_4{}^4 &= 0.\end{aligned}\right\} \quad . \quad . \quad . \quad (38{\cdot}48)$$

Hence

$$G = -\frac{6}{R^2}$$

and

$$G_1{}^1 - \tfrac{1}{2}\left(-\frac{6}{R^2} - 2\lambda\right) = 0,$$

since only $t_4{}^4$, among the components of **t**, is different from zero.

Consequently

$$-\frac{2}{R^2} + \frac{3}{R^2} + \lambda = 0$$

and

$$\boxed{\lambda = -\frac{1}{R^2}.} \quad . \quad . \quad . \quad . \quad . \quad (38{\cdot}481)$$

Further,

$$G_4{}^4 - \tfrac{1}{2}\left(-\frac{6}{R^2} + \frac{2}{R^2}\right) = \frac{8\pi\kappa\rho c^2}{c^4}$$

and consequently

$$\boxed{\frac{1}{R^2} = \frac{4\pi\kappa\rho}{c^2}.} \quad . \quad . \quad . \quad . \quad (38{\cdot}482)$$

Thus the radius of Einstein's world can be determined from the mean density, ρ, of the material spread through it. It is only possible to obtain a rather rough estimate of this density. It is obviously very small and compatible with an enormously big

value of R. H. Shapley estimates ρ to be 10^{-30} gm./cm.^{3} and when this is substituted in **(38·482)** we get

$$R = 3{\cdot}26 \times 10^{28} \text{ cm.} \quad . \quad . \quad . \quad . \quad . \quad \textbf{(38·49)}$$

or

$$R = 3{\cdot}45 \times 10^{10} \text{ light years}$$

which is approximately 10^{10} parsecs—a parsec is defined to be such a distance that the angle subtended at a point, situated at this distance from the sun, by the end points of that diameter of the earth's orbit which is at right angles to the line joining the sun and the point in question is one second. It is equal to 3·258 light years.

§ 38·5. The Volume and Mass of Einstein's World

Since

$$R^2 = x_1^2 + x_2^2 + x_3^2 + \xi^2,$$

it is permissible to adopt the angular co-ordinates χ, θ, and ϕ defined by

$$\left.\begin{aligned} x_1 &= R\cos\chi, \\ x_2 &= R\sin\chi\cos\theta, \\ x_3 &= R\sin\chi\sin\theta\cos\phi, \\ \xi &= R\sin\chi\sin\theta\sin\phi. \end{aligned}\right\} \quad . \quad . \quad . \quad \textbf{(38·5)}$$

In terms of these co-ordinates the spacial interval, $d\sigma$, is given by

$$d\sigma^2 = R^2\{d\chi^2 + \sin^2\chi(d\theta^2 + \sin^2\theta\, d\phi^2)\}, \quad . \quad \textbf{(38·51)}$$

and it will be seen that

$$\begin{aligned} g_{11} &= R^2, \\ g_{22} &= R^2\sin^2\chi, \\ g_{33} &= R^2\sin^2\chi\sin^2\theta, \end{aligned}$$

the remaining components, g_{mn}, vanishing. The associated value of $\sqrt{g}$—we are now dealing with space only—is

$$\sqrt{g} = R^3\sin^2\chi\sin\theta, \quad . \quad . \quad . \quad . \quad \textbf{(38·52)}$$

and for the element of volume, dV (cf. **§ 36·4**) we find

$$dV = R^3\sin^2\chi\sin\theta\, d\chi\, d\theta\, d\phi.$$

Finally, the total volume of the world is

$$V = R^3\int_0^{\pi}\int_0^{\pi}\int_0^{2\pi}\sin^2\chi\sin\theta\, d\chi\, d\theta\, d\phi,$$

or

$$\boxed{V = 2\pi^2R^3,} \quad . \quad . \quad . \quad . \quad . \quad \textbf{(38·53)}$$

and since it is occupied by matter of uniform density, ρ, the total mass in the world must be

$$M = 2\pi^2 R^3 \rho,$$

or, if we make use of **(38·482)**,

$$\boxed{M = \frac{\pi R c^2}{2\kappa},} \quad \ldots\ldots \textbf{(38·54)}$$

and we see that the total mass in Einstein's world is determined by its radius. The estimate of R in **(38·49)** yields

$$\boxed{\left\{\begin{matrix}\textbf{Total Mass in}\\ \textbf{Einstein's World}\end{matrix}\right\} = 6{\cdot}8 \times 10^{56} \textbf{ grams.}}$$

§ 38·6. de Sitter's Spherical World

Einstein's space-time is appropriately described as **cylindrical,** the axis of the cylinder representing the time. Its spacial part is spherical, i.e. the geometry of Einstein's space is of the same character as that of the surface of a sphere, or of the circumference of a circle. Hence Einstein's space-time is as the surface of a cylinder. It is of interest to compare it with another cosmological conception due to the Dutch astronomer de Sitter. de Sitter's world is (to a first approximation) an empty one and by contrast with Einstein's may be called **spherical** since the whole of his continuum of space plus time is spherical and not only the spacial part of it.

We shall naturally begin the study of it by using co-ordinates x_1, x_2, x_3, x_4, and ξ, connected by the relationship

$$R^2 = x_1^2 + x_2^2 + x_3^2 + x_4^2 + \xi^2, \quad \ldots \textbf{(38·6)}$$

where R is a constant—the radius of de Sitter's space-time world. The square of the interval is

$$ds^2 = dx_1^2 + dx_2^2 + dx_3^2 + dx_4^2 + d\xi^2.$$

As in **§ 38·4** we may write

$$\xi = \sqrt{R^2 - r^2},$$

where now

$$r^2 = x_1^2 + x_2^2 + x_3^2 + x_4^2.$$

Consequently

$$d\xi = \frac{-\,r\,dr}{\sqrt{R^2 - r^2}},$$

and thus

$$ds^2 = dx_1^2 + dx_2^2 + dx_3^2 + dx_4^2 + \frac{(r\,dr)^2}{R^2 - r^2}$$

and finally therefore

$$ds^2 = \left\{\delta_{mn} + \frac{x_m x_n}{R^2 - r^2}\right\} dx_m\, dx_n \quad . \quad . \quad (38{\cdot}61)$$

where $$\delta_{mn} = \begin{cases} 1, & m = n, \\ 0, & m \neq n, \end{cases}$$

and the summation convention is in operation.

Hence $$g_{mn} = \delta_{mn} + \frac{x_m x_n}{R^2 - r^2}$$

for all the g_{mn}. Further, the formulae **(38·47)** and **(38·471)** hold here for the derivatives of the g_{mn}, the only distinction between the investigation in **§ 38·4** and the present one is due to the fact that the formulae now apply for

$$m,\ n = 1,\ 2,\ 3,\ \text{and}\ 4.$$

When we evaluate the $G_m{}^p$ by the aid of **(38·472)** we easily find

$$G_1{}^1 = G_2{}^2 = G_3{}^3 = G_4{}^4 = -\frac{3}{R^2}.$$

Since $g_1{}^1 = g_2{}^2 = g_3{}^3 = g_4{}^4 = 1$ we have

$$G_1{}^1 = -\frac{3}{R^2} g_1{}^1,$$

$$G_2{}^2 = -\frac{3}{R^2} g_2{}^2,$$

and so on, and therefore in de Sitter's world

$$G_m{}^p = \lambda g_m{}^p \quad . \quad . \quad . \quad . \quad . \quad . \quad (38{\cdot}62)$$

where

$$\lambda = -3/R^2. \quad . \quad . \quad . \quad . \quad (38{\cdot}621)$$

It should be noted that R, the radius of de Sitter's world, is the radius of the whole space-time continuum and has not the same significance as the R of Einstein's world which is the radius of the spacial part of the world. It is clear that the R of de Sitter's world, like that of Einstein's, must be enormously great, since λ is bound to be very small.

It is instructive and helpful to study de Sitter's world in the following way: We imagine the spacial world to be finite, as is Einstein's world, for example, but instead of supposing the density of the matter in it to have the same constant value everywhere we shall imagine the whole of the stellar configurations to occupy only a minute portion of it; so that we shall be able to render an approximate account of it by treating the vast stellar universe as if it were an isolated particle. Let us now

introduce co-ordinates after the manner of § 37·1 and determine the g_{mn} as we did in that section, using, however (for empty space), the law

$$G_{mn} = \lambda g_{mn},$$

in which λ is a constant. We shall then approach de Sitter's world if we imagine the total mass in the world to approach the limit zero.

The equations **(37·17)** have now to be replaced by

$$\left.\begin{aligned}
G_{11} &\equiv \frac{N''}{2} - \frac{L'N'}{4} + \frac{N'^2}{4} - \frac{L'}{r} = \lambda e^L,\\
G_{22} &\equiv -e^{-L}\left\{r\left(\frac{L'-N'}{2}\right) - 1\right\} - 1 = \lambda r^2,\\
G_{33} &\equiv -\sin^2\theta\left[e^{-L}\left\{r\left(\frac{L'-N'}{2}\right) - 1\right\} + 1\right] = \lambda r^2\sin^2\theta,\\
G_{44} &= -c^2 e^{N-L}\left\{\frac{N''}{2} - \frac{L'N'}{4} + \frac{N'^2}{4} + \frac{N'}{r}\right\} = -c^2\lambda e^N.
\end{aligned}\right\} \quad (38\cdot63)$$

As in § 37·1 we find

$$N + L = \varepsilon,$$

where ε is a constant and therefore the interval **(37·12)** can be given the form:

$$ds^2 = e^L dr^2 + r^2\,d\theta^2 + r^2\sin^2\theta\,d\phi^2 - c^2 e^{-L}\{d(e^{\frac{\varepsilon}{2}}t)\}^2,$$

or, if we write t' for $e^{\frac{\varepsilon}{2}}t$,

$$ds^2 = e^L dr^2 + r^2\,d\theta^2 + r^2\sin^2\theta\,d\phi^2 - c^2 e^{-L}\,dt'^2. \quad . \quad (38\cdot64)$$

The constant, ε, has been eliminated by introducing a new unit for the time.

From the second (or third) of the equations **(38·63)** we get

$$-rL'e^{-L} + e^{-L} - 1 = \lambda r^2$$

and therefore

$$\frac{d}{dr}(re^{-L}) - 1 = \lambda r^2;$$

so that

$$re^{-L} - r = \frac{\lambda r^3}{3} + C,$$

where C is a constant of integration.

Hence

$$e^{-L} = 1 + \frac{C}{r} + \frac{\lambda r^2}{3},$$

and comparison with **(37·19)** and **(37·27)** makes it evident that

$$C = -2\kappa M/c^2,$$

where M is the mass of the universe, and we have finally

$$e^{-L} = 1 - \frac{2\kappa M}{c^2 r} + \frac{\lambda}{3} r^2. \quad . \quad . \quad . \quad . \quad (38{\cdot}65)$$

On passing to the limit of a practically empty world we get

$$ds^2 = \{1/(1 + \lambda r^2/3)\}dr^2 + r^2\, d\theta^2 + r^2 \sin^2\theta\, d\phi^2 - c^2(1 + \lambda r^2/3)dt^2, \quad . \quad . \quad (38{\cdot}66)$$

where dt has been written for dt', and, when we substitute $-3/R^2$ for λ (cf. **38·621**),

$$ds^2 = \{1/(1 - r^2/R^2)\}dr^2 + r^2\, d\theta^2 + r^2 \sin^2\theta\, d\phi^2 - c^2(1 - r^2/R^2)dt^2. \quad . \quad . \quad (38{\cdot}661)$$

This is an alternative expression for **(38·61).**

The geodesics in de Sitter's world are described by equations **(37·21)** if we substitute in these

$$1 - r^2/R^2 \text{ for } 1 - 2\kappa M/c^2 r,$$

and, just as in **§ 37·2,** we shall find it convenient to use the expression for the square, ds^2, of the interval instead of using the equation containing d^2r/ds^2. If we choose our reference system so that $d\theta/ds$ is initially zero while $\theta = \pi/2$, then, as in **§ 37·2,** $d\theta/ds$ will always be zero and θ will always be equal to $\pi/2$. If finally we take a geodesic for which the constant, $r^2\, d\phi/ds$ is zero, the expression for the square of the interval, ds, simplifies to

$$ds^2 = \frac{1}{\left(1 - \frac{r^2}{R^2}\right)} dr^2 - c^2\left(1 - \frac{r^2}{R^2}\right)dt^2 . \quad . \quad (38{\cdot}67)$$

Now the equation which corresponds to **(37·24)** is

$$\frac{dt}{ds} = \frac{K}{1 - \frac{r^2}{R^2}},$$

where K is a constant of integration. Hence on eliminating ds in **(38·67)** we get

$$\left(\frac{dr}{dt}\right)^2 = \left(1 - \frac{r^2}{R^2}\right)^2 \left\{ \frac{1 - \frac{r^2}{R^2}}{K^2} + c^2 \right\}. \quad . \quad . \quad (38{\cdot}671)$$

Let us consider the case where the velocity dr/dt is vanishingly small when $r = 0$. The last equation then gives

$$0 = \left\{ \frac{1}{K^2} + c^2 \right\},$$

so that for such a geodesic **(38·671)** becomes

$$\left(\frac{dr}{dt}\right)^2 = c^2\frac{r^2}{R^2}\left(1 - \frac{r^2}{R^2}\right)^2$$

and therefore

$$\frac{dr}{dt} = c\frac{r}{R}\left(1 - \frac{r^2}{R^2}\right). \quad . \quad . \quad . \quad . \quad . \quad \textbf{(38·68)}$$

The radial velocity dr/dt can easily be shown to be positive by inspecting the geodesic equation containing d^2r/ds^2.

We see from **(38·68)** that a body possessing a vanishingly small velocity in the neighbourhood of $r = 0$ will acquire an outward radial velocity in accordance with **(38·68).** When r approaches R this radial velocity is vanishingly small. Indeed, all velocities, that of light included, vanish at the outermost limits of de Sitter's world. So long, however, as $r^2 < < R^2$ **(38·68)** approximates to

$$\frac{dr}{dt} = c\frac{r}{R}. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(38·681)}$$

The actual world cannot of course be exactly like de Sitter's since the latter contains no mass; but it may possibly be a world with a quantity of matter in it which is insufficient to affect its large scale geometry appreciably, and this might indeed be that of de Sitter's world. Now we cannot doubt that R is enormously great, so that those distances, r, for which dr/dt in **(38·681)** is appreciable, may still be very large, though a small fraction of R. Now most of those very remote extra-galactic configurations known as the spiral nebulae have a velocity of recession which accords approximately with **(38·681).** There are exceptional instances of approaching nebulae; but we can explain this as due to the gravitational attraction of the masses which cause the actual world to differ locally from the ideal de Sitter one. The velocity of the spiral nebulae in the line of sight is inferred of course by ascribing the observed shift of the spectral lines of the light emitted by them to an associated Doppler effect. The displacement toward the red end which would result from **(38·681)** is easily worked out to be

$$\lambda_N - \lambda = \lambda r/R, \quad . \quad . \quad . \quad . \quad \textbf{(38·69)}$$

where λ is the wave-length of some spectral line in a terrestrial source and λ_N the wave-length of the same line in the light from the nebula.

The displacement of the spectral lines towards the red end of the spectrum is not wholly due to the velocity of recession. Considerations similar to those of **§ 37·5** indicate that atomic vibrations in such distant configurations as the spiral nebulae

must be slower and the wave-length of the emitted light longer in consequence, if we assume that the world approximates to de Sitter's conception of it. This is not a gravitational effect like that of **§ 37·5**; but the mathematical relationships that lead to it have the same form as those which led to the analogous effect described in that section.

To get an expression for this increase in wave-length we have to replace $\kappa M/c^2 r_s$ in **(37·5)** by $r^2/2R^2$. Thus we find

$$\lambda_N - \lambda = \frac{r^2}{2R^2}\lambda. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(38·691)}$$

When $r^2 < < R^2$, as we are assuming, the latter contribution to $\lambda_N - \lambda$ is negligible by comparison with that due to the recessional motion.

§ 38·7. The Expanding Universe

The cosmological speculations of Einstein and de Sitter rest on rather insecure foundations, but the latter find at least some support in the observations of the slight displacement of the spectral lines of the light from the spiral nebulae which are perhaps in as good agreement with **(38·69)** as their accuracy warrants. There is another way of regarding de Sitter's world which is implicit in the description of it given in the preceding section. We may write **(38·6)** in the form:

$$R_s^2 = R_e^2 + x_4^2, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(38·7)}$$

in which the radius, R, of de Sitter's world is represented by R_s and in which the sum $x_1^2 + x_2^2 + x_3^2 + \xi^2$ has been replaced by R_e^2 (cf. **38·41**). It will be observed that R_s is the radius of the space-time continuum and not the radius of the spacial world. This latter is represented by R_e. Now R_s is constant and x_4^2, when not zero, is negative. Hence R_e^2 and therefore also R_e increases as the *numerical* value of x_4^2 increases. Therefore R_e increases both when we go backwards in time from the instant $x_4 = 0$ and when we go forwards in time from this instant. Therefore before the instant $x_4 = 0$ de Sitter's world was contracting. At $x_4 = 0$ it reached its minimum radius and has since then been expanding. There is no such feature as expansion or contraction in Einstein's world. The formula **(38·7)** applies to it of course, but it is R_e which is constant and not R_s. Moreover in Einstein's world, as we have seen in **§ 38·4**, dx_4^2 is equal to $-c^2dt^2$, while in de Sitter's world dx_4^2 is equal to $-g_{44}c^2\,\mathrm{d}t^2$ where g_{44} is equal to $1 + x_4^2/(R_s^2 - r^2)$ **(38·21)**.

Einstein's world, in which, when we disregard trivial local irregularities, matter is uniformly distributed, is unstable and

recent cosmological speculation has tended in the direction of regarding the actual world as one which is passing out of a condition like that of Einstein's world towards the expanding world of de Sitter.

BIBLIOGRAPHY

The classical papers by LORENTZ, EINSTEIN, MINKOWSKI, and WEYL are published in a book called *Das Relativitätsprinzip* (Teubner, Leipzig). This has been translated by Perrett and Jeffery under the title *The Principle of Relativity* (Methuen).

E. CUNNINGHAM : The Principle of Relativity. (Cambridge.)
L. SILBERSTEIN : The Theory of Relativity. (Macmillan.)
M. LAUE : Das Relativitätsprinzip. (Vieweg.)
M. BORN : Einstein's Theory of Relativity. (Methuen.)
A. EDDINGTON : The Mathematical Theory of Relativity. (Cambridge.)
H. WEYL : Raum-Zeit-Materie. (Julius Springer.)
W. DE SITTER : On Einstein's Theory of Gravitation and its Astronomical Consequences. Monthly Notices R.A.S., 76, p. 699 ; 77, p. 155 ; 78, p. 3 (1916–17).
KALUZA : Zum Unitätsproblem der Physik. Preus. Akad. der Wissenschaften, Berlin, 54, p. 966 (1921).
W. WILSON : Relativity and Gravitation. Proc. Physical Soc., 31, p. 69 (1919); and Proc. Roy. Soc. A, 102, p. 478 (1922).

The following works of a popular or semi-popular character may be mentioned :

ALBERT EINSTEIN : The Theory of Relativity. (Methuen.)
—— The Meaning of Relativity. (Methuen.)
A. V. VASILIEV : Space Time Motion. (Chatto and Windus.)
A. EDDINGTON : Space, Time and Gravitation. (Cambridge.)

CHAPTER VIII

THE QUANTUM THEORY OF PLANCK

§ 38·8. The Nature of Heat

THE form of energy called **heat** has been defined in a practical way in **§ 15·5** and we associate with it two great principles : the first and second laws of thermodynamics which enable a rather wonderful co-ordination of thermal phenomena to be effected. These principles do not throw any light on the *nature* of heat : that is elucidated by the statistical theories—the kinetic theory of gases and statistical mechanics (cf. **§§ 11·8** to **13·5**)—which indicate that it is not a special form of energy, as is kinetic energy or electromagnetic energy, but simply energy of any or every form associated with **statistical equilibrium.** Briefly we use the term **heat** when we are dealing with an assemblage containing an enormous number of minute systems whose individual natures we may know or be able to infer, but which are so small and numerous that we are necessarily kept in ignorance of their individual energies, states of motion and configurations. The nature of statistical equilibrium is described in **§§ 12·9** to **13·4.** Each state of statistical equilibrium is characterized by a definite value of a quantity called the **modulus (§§ 13·1** and **13·3)** which we identify with **temperature.** The modulus has no meaning except in relation to a state of statistical equilibrium, and similarly temperature, in the strict sense of the term, has only a meaning when thermal equilibrium exists.

§ 38·9. Radiant Heat

Long ago (in the eighteenth century) it was observed that heat could be propagated by a process—to which the name **radiation** was given—very different from conduction or convection. The thermal energy propagated in this way was appropriately called **radiant heat.** It had the characteristic property associated with heat that its transference ceased with the establishment of temperature equality or uniformity. Later investigations of course revealed that radiant heat was physically

identical with light. This feature of radiant heat has its parallel in the identification of the heat content of a gas with the kinetic (or mechanical) energy of its constituent molecules. Radiant heat, it was discovered, travelled through free space with the same velocity as light, could be polarized, and so on. The only difference that could be detected between it and light was the purely subjective one that hot bodies were often observed to radiate without exciting the sensation of sight.

The fact of the possibility of temperature equilibrium being associated with radiation led Prévost in 1791 to his theory of exchanges according to which all bodies, hot or cold, are radiating heat; uniform temperature resulting when they are absorbing radiant heat at the same rate as they are emitting it. The laws discovered by G. Kirchhoff (1859) are really amplifications of Prévost's theory of exchanges. The possibility of temperature equilibrium necessitates that a good emitter of radiant heat must also be a good absorber and a good reflector a bad absorber. These laws of Kirchhoff hold not only for the totality of the radiant heat, but also in detail for its components. Good absorbers of light of a particular wave-length must also be good emitters of light of that particular wave-length. Tourmaline, for example, strongly absorbs light or radiant heat polarized in a certain plane and when heated emits correspondingly copiously radiant heat which is polarized in just that particular way. These old discoveries about the nature of radiant heat were associated with the further discovery that the physical thing called **light** is not limited to something which excites the sensation of sight: that the spectrum in fact extends out beyond both visible limits.

§ 39. Full Radiation

A state of thermal (i.e. statistical) equilibrium must exist, or will gradually be established, in any region contained within a closed surface which is maintained at a uniform temperature. This is equally true whether the region is occupied by some material such as a gas, for example, or is vacuous.[1] And when equilibrium is established every point in the region has the same temperature, namely, the temperature of the surrounding surface. Moreover, when the material filling the enclosure is the same throughout it, not only is the temperature uniform everywhere in it, but the state of affairs in other respects, e.g. energy

[1] Vide **§ 34.** We must suppose the 'vacuous' region to be occupied by electromagnetic waves or, alternatively, by photons.

density, is identical at all points and independent of the material of the enclosing wall.

If the volume of a given quantity of a material medium be altered, the state of stress in it must be changed consequently and with it the energy density, which therefore depends not only on the temperature, but also on the volume. In the case of the vacuous enclosure, however, the character of the medium is not affected by volume changes and the **energy density is determined completely by the temperature of the enclosing wall.** How different this is from the behaviour of a material medium will be realized when it is remembered that the **whole energy** of a gas—**not its energy density**—is determined (practically) by the temperature of the wall of the vessel containing it. We may say, in fact, that the state of affairs at any point of the vacuous enclosure is independent of the volume for a given temperature of the enclosing wall. It is, in fact, a function of the temperature only. The independence of the state of affairs within the enclosure on the nature of the material of its wall enables us to assert that the radiant heat in it has precisely the same character as that which it would have if the enclosing surface were **black,** this term being used for a surface which absorbs completely all the radiation incident on it. We usually speak of the radiation within the enclosure as **full radiation.** It is perhaps hardly so good a term as **radiant heat.** Like any other form of heat, it is essentially an equilibrium phenomenon, and if we were to imagine an aperture in the wall of the enclosure the emerging radiation would not necessarily be full radiation since the presence of the aperture, together with a lower external temperature, for example, would make equilibrium impossible. When, however, the aperture is sufficiently small compared with the dimensions of the enclosure, the departure from equilibrium will be negligible and the emerging radiation will have practically the character of that which would be emitted by a perfectly black surface. Indeed, the surface bounded by the periphery of the aperture will simulate a black surface almost perfectly. Any radiation that may enter the enclosure through it can, practically speaking, never emerge again. It is in fact completely absorbed. The experimental devices for producing black body (or full) radiation for purposes of observation and measurement are based on these considerations.

§ 39·05. The Laws of Full Radiation

A very simple application of thermodynamics (cf. **§ 39·1**) leads to the law named after Stefan (1879) and Boltzmann (1884),

according to which the energy per unit volume in the enclosure is proportional to the fourth power of the absolute temperature expressed in terms of Kelvin's work scale.

A much more formidable problem is that of rendering a theoretical account of the distribution of the energy in the spectrum of the full radiation. The nature of this problem may be described by the aid of a diagram. In Fig. 39·05 the function, e, which we shall use to express the spectral distribution of energy, is plotted against the wave-length, λ, for a given temperature. The full line represents e as a function of λ at the temperature, T; the broken line represents e as a function of λ at the higher temperature, T'. The function e is so defined that the area bounded by any two ordinates and the curve itself, e.g.

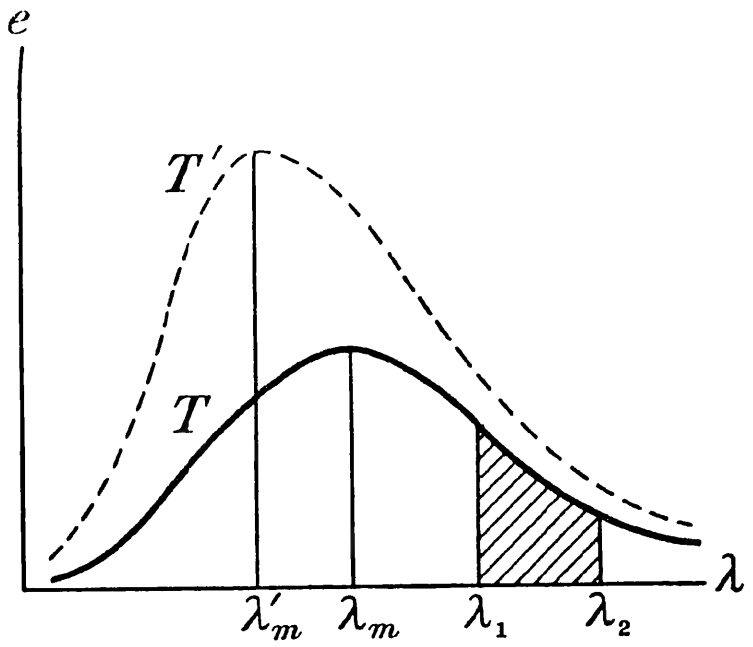

FIG. 39·05

the shaded area between the ordinates at λ_1 and λ_2 is equal to the energy per unit volume of all the vibrations whose frequencies lie between the limits corresponding to those two wave-lengths.

The chief problem of the theory of full radiation is that of expressing e as a function of λ and T, or of the frequency, ν and T. Its special interest is due to the fact that its solution by Max Planck in 1900 (cf. **§ 39·9**) necessitated that departure from the methods and principles of classical physical theory —and we must include relativity in this category—with which we are now familiar under the names **quantum theory** and **quantum dynamics.**

By an interesting application of thermodynamics W. Wien **(§ 39·4)** was able to prove that the function, e, must have the form

$$e = \lambda^{-5} f(\xi), \quad . \quad . \quad . \quad . \quad . \quad \textbf{(39·05)}$$

where ξ means the product λT and f is a function whose form is not disclosed by Wien's method. The laws of Stefan and Boltzmann and of Wien contain all that can be achieved by the methods of classical physical theory in connexion with the problems of full radiation, as we shall understand better when we come to deal with them in detail; but Wien made a further important contribution towards the final solution of the distribution problem in the following formula, obtained, as we now believe, by an unsound method:

$$e = A\lambda^{-5}e^{-B/\lambda T} \quad [1] \quad . \quad . \quad . \quad . \quad \textbf{(39·06)}$$

in which A and B are universal positive constants. This formula fits the results of observation quite well when the product λT is sufficiently small and it is in agreement with the perfectly correct formula **(39·05)**. In fact, it gives the function f in this formula the shape

$$Ae^{-B/\delta T}.$$

The spectral distribution problem was also attacked by Lord Rayleigh, who obtained a result equivalent to

$$e = 8\pi kT/\lambda^4, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(39·07)}$$

where k means the entropy constant of **§ 13·4**. This result is also in agreement with **(39·05)** since we may give it the form

$$e = 8\pi k(\lambda T)/\lambda^5.$$

Now Rayleigh's formula **(39·07)** was obtained by a perfectly sound method and we are forced to ascribe its failure to agree with the observational results, as Planck did, to the inadequacy of classical dynamical principles, more especially the principle of equipartition (cf. **§§ 12** and **13·5**) on which its deduction was based. Rayleigh's formula fits the observations however when the product, λT, is sufficiently great. This fact illustrates a familiar feature of classical theory: namely, its adequacy in macrophysical applications.

The truth is obviously contained in some formula which approaches **(39·06)** and **(39·07)** in the limiting cases respectively of very small and very large values of λT. Such a formula is the following one which we get by making a slight change in **(39·06)**:

$$e = A\lambda^{-5}/(e^{B/\lambda T} - 1). \quad . \quad . \quad . \quad . \quad \textbf{(39·08)}$$

It obviously becomes identical with Wien's formula in one limiting case, since the subtraction of unity from the exponential is

[1] It is unfortunate that e has two senses in this equation; but one of the e's is so obviously the exponential e that no confusion, it may be hoped, will arise.

of no consequence when λT is very small. In the other limiting case it becomes

$$e = A\lambda^{-5}/(B/\lambda T) = (A/B)(T/\lambda^4)$$

so that

$$A/B = 8\pi k. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(39·09)}$$

The formula **(39·08)** fits the observations exceedingly well and there is no doubt it is the right one. It was discovered by Planck, who succeeded in giving it a theoretical basis. This will be studied in **§ 39·9.** Meanwhile we might note that the exponential in Wien's formula **(39·06)** is suggestive of the exponential in Maxwell's distribution formula **(12·16)** and even more so of the exponential

$$e^{-\beta E} = e^{-E/\Theta} = e^{-E/kT}$$

in **(13·171).** And when we remember that many phenomena, notably those of photo-electricity, force us to regard light (cf. **§§ 34·6** and **37·4**) as having a corpuscular constitution we are tempted to look upon full radiation as a sort of gas, a **photon gas.** We are thus led to identify the exponents $B/\lambda T$ and E/kT, where E means the energy of a photon. Thus

$$E = Bk/\lambda, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(39·091)}$$

or

$$E = Bk/c\tau, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(39·092)}$$

where c is the velocity of radiation in free space. Thus we may write for the energy of a photon (cf. **§ 41·2** et seq.)

$$E = h/\tau = h\nu, \quad . \quad . \quad . \quad . \quad \textbf{(39·093)}$$

where

$$h = Bk/c,$$

or

$$B = ch/k. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(39·094)}$$

The expression **(39·093)** for the energy of a photon is suggested also by the analogy or correspondence between dynamics and geometrical optics briefly discussed in **§ 9·4** and much more fully in **§ 41·5.**

From equations **(39·09)** and **(39·094)** we have

$$A = 8\pi ch. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(39·095)}$$

The substitution of these expressions for A and B in Planck's formula **(39·08)** gives it the familiar shape:

$$e = 8\pi ch\lambda^{-5}/(e^{\frac{ch}{k\lambda T}} - 1). \quad . \quad . \quad . \quad \textbf{(39·096)}$$

Now $ed\lambda$ means the quantity of radiant energy per unit volume associated with wave-lengths in the narrow range between λ and $\lambda + d\lambda$. We may write it as $-uv\,dv$ where dv is the range of

frequencies corresponding to $d\lambda$ and since $\lambda = c/\nu$ and consequently $d\lambda = -c\,d\nu/\nu^2$ we get from

$$e\,d\lambda = -u_\nu\,d\nu$$

$$u_\nu = 8\pi h\nu^3/c^3(e^{\frac{h\nu}{kT}} - 1) \quad . \quad . \quad . \quad \textbf{(39·097)}$$

which is an alternative form of equation **(39·096).**

The methods by which we have reached this law of Planck in this section are largely empirical; but they contain already suggestions for a new and rather revolutionary dynamical theory which will occupy us in the chapters that follow.

§ 39·1. The Law of Stefan and Boltzmann

Let us imagine the enclosure of **§ 39** to undergo a reversible expansion. We may then apply to it the thermodynamical formula **(15·91)**

$$dQ = dU + p\,dV, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(39·1)}$$

where dQ means the heat supplied, U is the radiant energy in the enclosure, p is the pressure it exerts on the containing wall and dV is the increment in volume associated with dQ and dU.

Now

$$U = uV, \quad . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(39·11)}$$

where u means the energy per unit volume, and

$$p = u/3 \quad . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(39·12)}$$

as shown in **§ 27.** Therefore

$$dQ = V\,du + (u + p)dV,$$

and since u is a function of T only (cf. **§ 38·9**),

$$dQ = V\frac{du}{dT}dT + \frac{4u}{3}dV.$$

On dividing both sides of this equation by the temperature we obtain an expression for the associated increment in entropy, namely

$$d\phi = \frac{V}{T}\frac{du}{dT}dT + \frac{4u}{3T}dV, \quad . \quad . \quad . \quad \textbf{(39·13)}$$

and hence, by the thermodynamical rule **(16·91),**

$$\frac{\partial}{\partial V_T}\left\{\frac{V}{T}\frac{du}{dT}\right\} = \frac{\partial}{\partial T_v}\left\{\frac{4u}{3T}\right\}.$$

Therefore

$$\frac{1}{T}\frac{du}{dT} = \frac{4}{3T}\frac{du}{dT} - \frac{4u}{3T^2}$$

and so

$$4\frac{dT}{T} = \frac{du}{u}.$$

Therefore

$$\boxed{u = aT^4,} \qquad . \quad . \quad . \quad . \quad \textbf{(39·14)}$$

where a is a constant which we may term the **Stefan-Boltzmann constant** or the **constant of full radiation.**

§ 39·2. Alternative Statements of the Law of Stefan and Boltzmann

The quantity of radiant energy, per unit volume, travelling in directions contained within, or defined by, a narrow solid angle, $d\Omega$ is clearly

$$ud\Omega/4\pi. \quad . \quad . \quad \textbf{(39·2)}$$

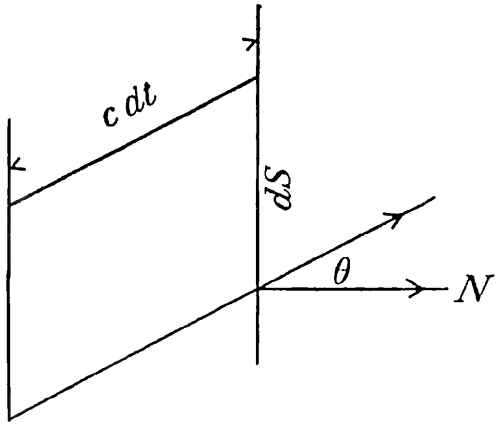

Fig. 39·2

To get the quantity passing across a small surface dS during the time dt and travelling in these directions we have to multiply the expression **(39·2)** by the volume of the cylinder shown in Fig. 39·2, namely $c\,dt\,dS\cos\theta$, where θ is the angle between the direction of the axis of $d\Omega$ and that of the normal, N, to the surface dS. We thus obtain

$$\frac{uc}{4\pi}\sin\theta\cos\theta\,d\theta\,d\phi\,dt\,dS, \quad . \quad . \quad . \quad \textbf{(39·21)}$$

since $d\Omega = \sin\theta\,d\theta\,d\phi$, where ϕ is the azimuthal angle.

On integrating with respect to θ and ϕ we obtain

$$\frac{uc}{4}dS\,dt \quad . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(39·22)}$$

for total quantity of energy crossing dS from one side to the other during the time dt and

$$uc/4 \quad . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(39·221)}$$

for the quantity reckoned per unit area per unit time.

The quantity $uc/4\pi$ in **(39·21)** may be written

$$\frac{uc}{4\pi} = \frac{ac}{4\pi}T^4 = \sigma' T^4,$$

and $uc/4$ in **(39·22)** may be written

$$\frac{uc}{4} = \frac{ac}{4}T^4 = \sigma T^4.$$

The three constants, a, σ, and σ', are related by the equations

$$\left.\begin{aligned} \sigma &= ac/4, \\ \sigma' &= ac/4\pi. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad \textbf{(39·23)}$$

We may of course call any one of these constants, a, σ, σ', the **constant of full radiation.** Their experimentally determined values are

$$\begin{aligned} a &= 7{\cdot}67 \times 10^{-15} \\ \sigma &= 5{\cdot}75 \times 10^{-5} \\ \sigma' &= 1{\cdot}83 \times 10^{-5} \end{aligned}$$

the energy density, u, being expressed in ergs per cubic centimetre.

§ 39·3. Entropy of Full Radiation

If we replace u in **(39·13)** by aT^4 we get for $d\phi$, the increment of the entropy of the full radiation,

$$d\phi = 4aVT^2\,dT + \frac{4}{3}aT^3\,dV,$$

and thus we obtain for ϕ, if we make the constant of integration zero,

$$\phi = \frac{4}{3}aT^3V. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(39·3)}$$

This expresses the entropy associated with the radiation in the enclosure.

§ 39·4. Wien's Displacement Law

During any reversible change or process a thermodynamical system remains in equilibrium and if the radiation in a vacuous enclosure is involved in such a process it must necessarily remain full radiation. Let us consider a reversible adiabatic expansion. The entropy is bound to remain constant and consequently by **(39·3)**

$$T^3V = \text{constant}. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(39·4)}$$

This is the relation between the temperature and volume of full radiation undergoing a reversible adiabatic change and for a small volume increment

$$dV_\phi = -(3V/T)dT_\phi, \quad . \quad . \quad . \quad . \quad \textbf{(39·41)}$$

the subscript, ϕ, reminding us that the increments dV_ϕ and dT_ϕ are those associated with constant entropy.

Let us now consider the radiation, between the frequency limits

$$\nu \text{——} \nu + \Delta\nu,$$

which is travelling in directions contained within the limits of a narrow solid angle, $d\Omega$ (cf. **§ 39·2**), and incident at an angle, θ, on a small portion, dS, of the wall of the enclosure. In order to have a short name for it we may call it the radiation in the region $(\Delta\nu, d\Omega, \theta)$. We shall suppose dS to be perfectly reflecting [1] and, in consequence of the reversible adiabatic expansion, to be moving outwards with a minute normal velocity, v. The energy from the region $(\Delta\nu, d\Omega, \theta)$ which reaches dS in the time dt will obviously be equal to

$$\Delta\nu . \frac{u_\nu \, d\Omega}{4\pi} . c \, dt \, dS \cos\theta,$$

(cf. **§ 39·2**), if we define u_ν so that $u_\nu \Delta\nu$ is the energy per unit volume associated with the vibrations in the frequency range $\nu \text{——} \nu + \Delta\nu$. If we replace $d\Omega$ by $\sin\theta \, d\theta \, d\phi$ we see that the energy from the region $(\Delta\nu, d\Omega, \theta)$ which reaches dS in the time dt is

$$\Delta\nu . \frac{u_\nu c}{4\pi} . \sin\theta \cos\theta \, d\theta \, d\phi \, dS \, dt. \quad . \quad . \quad . \quad \textbf{(39·42)}$$

After reflexion at dS this energy necessarily leaves the region $(\Delta\nu, d\Omega, \theta)$. Some energy of course will enter this region by reflexion at dS and this must necessarily have, before reflexion, a slightly higher frequency, namely

$$\nu' = \nu\left(1 + \frac{2v \cos\theta}{c}\right), \quad . \quad . \quad . \quad . \quad \textbf{(39·43)}$$

in order that the Doppler effect due to the outward motion of dS with the velocity, v, may bring it into the frequency range $\nu \text{——} \nu + \Delta\nu$. Clearly the energy entering the region $(\Delta\nu, d\Omega, dS)$ by reflexion at dS during the time dt will be

$$\frac{\Delta\nu . \left\{u_\nu + \dfrac{\partial u_\nu}{\partial\nu} . \dfrac{2v\nu \cos\theta}{c}\right\} c}{4\pi} . \sin\theta \cos\theta \, d\theta \, d\phi \, dS \, dt.$$

The gain in energy of the region $(\Delta\nu, d\Omega, \theta)$ by reflexion at dS

[1] This assumption is not unlike that made in **§ 11·9** in deducing Boyle's law and we might avoid it altogether by a device similar to the one adopted there.

during the time dt is obtained by subtracting **(39·42)** from this expression. We thus get

$$\frac{\Delta\nu}{2\pi}.\nu\frac{\partial u_\nu}{\partial\nu}.\sin\theta\cos^2\theta\,d\theta\,d\phi.dV, \quad . \quad . \quad . \quad \textbf{(39·44)}$$

where $$dV = v\,dt\,dS$$

is the increment in volume associated with the outward motion of dS during the time interval dt.

To get an expression for the total gain in energy in the frequency range from ν to $\nu+\Delta\nu$ due to the expansion dV we must integrate with respect to θ and ϕ, the limits being respectively

$$0\text{——}\pi/2 \text{ and } 0\text{——}2\pi.$$

In this way we obtain

$$\Delta\nu.\frac{\nu}{3}\left(\frac{\partial u_\nu}{\partial\nu}\right)_T dV_\phi. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(39·45)}$$

The subscript, T, is used to indicate that the differentiation is a partial one with respect to ν, the temperature, T, being the other variable on which u_ν depends, is not altered in the differentiation.

Now the increment **(39·45)** evidently represents precisely the same thing as

$$\Delta\nu.d(u_\nu V)_\phi,$$

or $$\Delta\nu.\{V(du_\nu)_\phi + u_\nu\,dV_\phi\},$$

which may be written

$$\Delta\nu\left\{V\left(\frac{\partial u_\nu}{\partial T}\right)_\nu dT_\phi + u_\nu\,dV_\phi\right\}.$$

On equating this to **(39·45)** we have

$$\frac{\nu}{3}\left(\frac{\partial u_\nu}{\partial\nu}\right)_T dV_\phi = V\left(\frac{\partial u_\nu}{\partial T}\right)_\nu dT_\phi + u_\nu\,dV_\phi,$$

and on replacing dV_ϕ by the expression **(39·41)** we obtain the differential equation

$$3u_\nu = \nu\left(\frac{\partial u_\nu}{\partial\nu}\right)_T + T\left(\frac{\partial u_\nu}{\partial T}\right)_\nu, \quad . \quad . \quad . \quad \textbf{(39·46)}$$

which contains **Wien's displacement law.** The factor, 3, on the left suggests the convenience of introducing a new function, ψ, defined by

$$\psi(\nu, T) \equiv u_\nu/\nu^3. \quad . \quad . \quad . \quad . \quad \textbf{(39·47)}$$

On substitution in **(39·46)** we get

$$\nu\left(\frac{\partial\psi}{\partial\nu}\right)_T + T\left(\frac{\partial\psi}{\partial T}\right)_\nu = 0, \quad . \quad . \quad . \quad . \quad \textbf{(39·48)}$$

which is a simpler equation than **(39·46)**.

Since ψ is some function of the independent variables ν and T,

$$d\psi = \left(\frac{\partial\psi}{\partial\nu}\right)_T d\nu + \left(\frac{\partial\psi}{\partial T}\right)_\nu dT$$

and consequently

$$0 = \left(\frac{\partial\psi}{\partial\nu}\right)_T d\nu_\psi + \left(\frac{\partial\psi}{\partial T}\right)_\nu dT_\psi,$$

or

$$0 = \left(\frac{\partial\psi}{\partial\nu}\right)_T \left(\frac{\partial\nu}{\partial T}\right)_\psi + \left(\frac{\partial\psi}{\partial T}\right)_\nu \quad . \quad . \quad . \quad \textbf{(39·481)}$$

On eliminating $(\partial\psi/\partial T)_\nu$ from **(39·48)** and **(39·481)** we get

$$\nu = T\left(\frac{\partial\nu}{\partial T}\right)_\psi.$$

Thus when ψ is kept constant

$$\frac{d\nu}{\nu} = \frac{dT}{T};$$

or $T/\nu = \text{constant}$

when $\psi = \text{constant}.$

This means that $T/\nu = \text{function } (\psi),$

or, what amounts to the same thing,

$$\psi = \psi(\eta)$$

where $\eta \equiv T/\nu.$

Finally, therefore,

$$\boxed{u_\nu = \nu^3\psi(\eta)} \quad . \quad . \quad . \quad . \quad . \quad \textbf{(39·49)}$$

This is one of the expressions of **Wien's displacement law.**

Now $u_\nu\, d\nu$ means the energy per unit volume associated with the range of frequencies ν——$\nu + d\nu$. Therefore

$$-u_\nu\, d\nu = e\, d\lambda,$$

(cf. **§ 39·05**) if $d\lambda$ is the differential of the wave-length corresponding to $d\nu$. Now

$$d\nu = -c\, d\lambda/\lambda^2$$

and consequently

$$u_\nu c/\lambda^2 = e. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(39·491)}$$

On substituting in **(39·49)** we get

$$\lambda^2 e/c = c^3\psi(\lambda T/c)/\lambda^3,$$

or

$$\boxed{e = \lambda^{-5} f(\xi)}$$

where $\xi = \lambda T$ and f is a, so far, undetermined function. This is Wien's law in the form **(39·05)** already given.

§ 39·5. Properties of the Maxima of the Spectral Distribution Curves

The last formula may be written

$$e = T^5 f(\xi)/\xi^5$$

or

$$e = T^5 \phi(\xi), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (39{\cdot}5)$$

where $\phi(\xi) \equiv f(\xi)/\xi^5$, which expresses Wien's law in still another way.

If now we keep T constant or, in other words, if we keep to the energy distribution curve corresponding to the temperature, T, (Fig. 39·05) and differentiate with respect to λ we shall get

$$de = T^6 \phi'(\xi) d\lambda,$$

where

$$\phi'(\xi) \equiv d\phi(\xi)/d\xi,$$

so that the maximum of the curve is determined by

$$\phi'(\xi) = 0. \quad . \quad . \quad . \quad . \quad . \quad . \quad (39{\cdot}51)$$

This last equation it will be noticed yields a definite numerical value, which we may call ξ_m, for ξ which is quite independent of the temperature. The product $\lambda_m T$ (cf. Fig. 39·05), for example, is equal to the product $\lambda_m' T'$. Experimental observations confirm this and yield

$$\xi_m = \lambda_m T = 0{\cdot}288 \quad . \quad . \quad . \quad . \quad (39{\cdot}52)$$

approximately, λ_m being expressed in centimetres.

Planck's law **(39·096),** for example, gives the function, f, of **(39·05)** the form:

$$8\pi ch(e^\beta - 1)^{-1}$$

where

$$\beta \equiv ch/k\lambda T = ch/k\xi, \quad . \quad . \quad . \quad . \quad (39{\cdot}53)$$

so that the function, ϕ, of **(39·5)** is

$$\phi(\xi) \equiv 8\pi ch\xi^{-5}(e^\beta - 1)^{-1}. \quad . \quad . \quad . \quad (39{\cdot}55)$$

Therefore equation **39·51** becomes:

$$\beta_m = 5(1 - e^{-\beta_m}). \quad . \quad . \quad . \quad . \quad (39{\cdot}56)$$

It is easy to solve this equation by successive approximations. To begin with, it is satisfied to a rather rough approximation by $\beta_m = 5$. If we substitute this in the exponential, which is small compared with unity, we get

$$\beta_m = 5(1 - e^{-5}),$$

which yields the better approximation, $\beta_m = 4{\cdot}9665$. To four decimal places we find, on continuing this process

$$\beta_m = ch/k\lambda_m T = 4{\cdot}9651 \ldots \quad . \quad . \quad (39{\cdot}57)$$

The subscript, m, is to remind us that we are dealing with values of β and λ appropriate to a maximum. With the aid of equation **(39·57)** we can calculate the ratio h/k from the experimentally determined value of $\lambda_m T$ and the known values of c and the number β_m.

On substituting the constant value, $\lambda_m T$, for ξ in **(39·5)** we learn that the values of e at the maxima of the energy distribution curves are proportional to the 5th power of the temperature. In fact $\phi(\xi_m)$ is a constant.

§ 39·6. Evaluation of the Constants of Boltzmann and Planck

From **(39·097)** we obtain the following expression for u, the energy per unit volume in the vacuous enclosure of **§ 39,**

$$u = \frac{8\pi h}{c^3}\int_0^\infty \frac{\nu^3 d\nu}{e^\beta - 1}, \quad . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(39·6)}$$

in which β means $h\nu/kT$. It is convenient to give it the form

$$u = \frac{8\pi k^4 T^4}{c^3 h^3}\int_0^\infty \frac{\beta^3 d\beta}{e^\beta - 1}, \quad . \quad . \quad . \quad . \quad \textbf{(39·61)}$$

which reveals that Planck's formula is in agreement with that of Stefan and Boltzmann. The integral can be expanded as follows:

$$\int_0^\infty \beta^3 d\beta \{e^{-\beta} + e^{-2\beta} + e^{-3\beta} + \ldots\},$$

and its numerical value is therefore

$$6 \times \left(\frac{1}{1^4} + \frac{1}{2^4} + \frac{1}{3^4} + \ldots\right) = 6 \times 1{\cdot}0823 \ldots$$

Since $u = aT^4$, where $a = 7{\cdot}67 \times 10^{-15}$ we have the numerical equation

$$7{\cdot}67 \times 10^{-15} = \frac{8\pi \times 6 \times 1{\cdot}0823}{c^3}\left(\frac{k^4}{h^3}\right),$$

from which we get

$$\frac{k^4}{h^3} = 1{\cdot}27 \times 10^{15} \quad . \quad . \quad . \quad . \quad . \quad \textbf{(39·62)}$$

On the other hand the observational value of $\lambda_m T$, which we may term **Wien's constant,** is 0·288 and **(39·57)** gives us

$$\frac{k}{h} = \frac{c}{0{\cdot}288 \times 4{\cdot}9651}, \quad . \quad . \quad . \quad . \quad \textbf{(39·63)}$$

and by combining **(39·62)** and **(39·63)** we obtain the following numerical values for h and k

$$\left.\begin{aligned} h &= 6{\cdot}55 \times 10^{-27} \text{ ergs} \times \text{sec.} \\ k &= 1{\cdot}37 \times 10^{-16} \text{ erg/°C.} \end{aligned}\right\} \quad . \quad . \quad . \quad \textbf{(39·64)}$$

§ 39·7. Loschmidt's Number and the Elementary Ionic Charge

Planck determined the values of h and k in this way and he made use of the value he found for k, namely $1{\cdot}34 \times 10^{-16}$ ergs per °C., to obtain an estimate of **Loschmidt's number** (cf. **§§ 12·6** and **12·8**). The number obtained by Planck was $6{\cdot}2 \times 10^{23}$, the first reasonably precise estimate of Loschmidt's number. And dividing the charge, $2{\cdot}896 \times 10^{14}$ ordinary electrostatic units, conveyed by a gram atom of hydrogen in electrolysis by his estimate of Loschmidt's number he was able further to calculate the value $4{\cdot}67 \times 10^{-10}$ O.E.S. units for the elementary ionic charge. This result was obtained by Planck at a time when the more direct methods of determining this quantity were only just beginning to be applied and years before Millikan used his well-known method. It was a very remarkable achievement and it constituted a confirmation of the correctness of his radiation formula and therefore of his new theory since his value of k was reached by the use of his new formula in the manner just described and the later measurements of the ionic charge by Millikan yielded a result very near to that calculated by Planck. Of course the value of k and consequently also Loschmidt's number and the ionic charge may be obtained from measurements carried out on full radiation and confined to spectral regions where λ is sufficiently great, or rather to spectral regions where λT is sufficiently great, since under these circumstances Rayleigh's formula **(39·03)** applies and it contains the single universal constant, k. So that full radiation measurements enable these quantities to be determined—if perhaps with less precision—without involving any departure from classical principles.

The fact that we can determine Loschmidt's number from one of the constants of the spectral distribution formula does not of itself justify us in suspecting full radiation to have a *quasi* gas constitution. The significance of the constant, k, as we

learned in §§ **13·4** and **13·5,** is much wider than is indicated by its rôle as gas constant per molecule. From the point of view of the classical theory it might be described as the constant which fixes the value of the average kinetic energy per degree of freedom of the individual systems in an assemblage which is in temperature equilibrium, and it does not matter whether the individual systems are molecules, vibrations in a more or less continuous medium, or what.

§ 39·8. The Classical Theory of Full Radiation

To lead the way to a theoretical basis for **Planck's formula (39·096)** it is best to study the problem of full radiation firstly from the point of view of Newtonian dynamical principles. We shall revert temporarily to the aether hypothesis and approach the problem of the spectral distribution of the energy in full radiation by considering a vacuous enclosure in the shape of a cube, each edge of which is of unit length. We may imagine the state of motion of the aether in it to be subjected to a Fourier analysis, i.e. we may imagine it to be a superposition of simple harmonic vibrations constituting stationary waves with nodal planes which divide the edges of the cube into integral numbers of intervals. This can be illustrated by the rather simpler example of a cord stretched between two points to which its ends are fixed. Any state of motion of the cord can be regarded as a superposition of stationary waves dividing it into intervals each equal in length to $\lambda/2$ **(§ 9).** When the cord is divided into n intervals the associated frequency of vibration, ν, is

$$\nu = nV/2L, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(39·8)}$$

where V is the (phase) velocity of a wave of such frequency along the cord and L is the length of the latter. If we confine our attention to motions in one plane there is *one* state of simple harmonic vibration for each value of n and the number of independent states of vibration with frequencies between zero and ν is in consequence of **(39·8)**

$$n = 2L\nu/V,$$

and when n is a very large integer, i.e. when ν is very large, we may write for the number of vibrations between ν and $\nu + d\nu$,

$$\frac{2L}{V}d\nu.$$

This refers to motions in a certain plane only. To remove this limitation we must multiply by 2, since the most general motion of the cord is a superposition of motions in some plane and quite

independent motions perpendicular to it. Hence we obtain for the total number of vibrations with frequencies between the limits ν and $\nu + d\nu$

$$\frac{4L}{V}d\nu.$$

This illustration will guide us in dealing with the enclosure mentioned above. A particular mode of simple harmonic vibration will divide the edge of the cube which is parallel to the X axis of a suitably placed system of rectangular co-ordinates into n_1 intervals, that parallel to the Y axis into n_2 intervals, and the remaining edge into n_3 intervals. Since each edge of the cube is of unit length the intervals are respectively equal in length to $1/n_1$, $1/n_2$, and $1/n_3$. If now the normal to a set of nodal planes makes the angles θ_1, θ_2, and θ_3 respectively with the X, Y, and Z axes

$$(1/n_1)\cos\theta_1 = \lambda/2$$

or

$$\cos\theta_1 = n_1\lambda/2.$$

Similarly

$$\cos\theta_2 = n_2\lambda/2,$$
$$\cos\theta_3 = n_3\lambda/2.$$

Hence

$$(n_1^2 + n_2^2 + n_3^2)\lambda^2/4 = 1,$$

or, since $\nu\lambda = c$,

$$n_1^2 + n_2^2 + n_3^2 = 4\nu^2/c^2. \quad . \quad . \quad . \quad \textbf{(39·81)}$$

It will be remembered that, in the illustration just given of the stretched cord, a particular value of the integer, n, was associated with *two* independent states of vibration, since any state of motion of the cord is a superposition of motions in any plane containing it and perfectly independent motions at right angles to this plane. So in the case under consideration any set of values of n_1, n_2, and n_3 is associated with *two* independent vibrations, one associated with movements parallel to any plane containing the normal to the nodal planes and the other with movements perpendicular to it. Let us represent the various sets of values of n_1, n_2, and n_3 in a Euclidean space, using rectangular axes of co-ordinates parallel to the edges of the vacuous

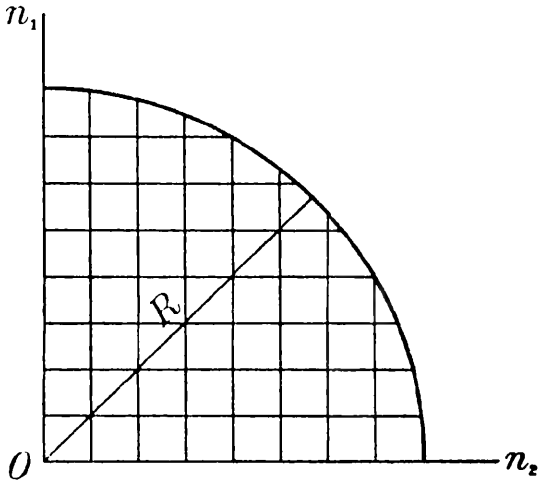

FIG. 39·8

cubical enclosure. We measure off a distance along one axis to represent n_1, along another axis to represent n_2, and so on. For each set of values n_1, n_2, n_3 there will be a corresponding point in the representative space (Fig. 39·8). The points representing all frequencies from zero to some upper limit, ν, will (cf. **39·81**), it is clear, be contained in an octant of a sphere of radius, R, and

$$R^2 = n_1^2 + n_2^2 + n_3^2.$$

Moreover, each point is situated at one corner of a unit cube and the total number of points associated with the range of frequencies mentioned will be identical with the volume of the octant, namely

$$\pi R^3/6 \text{ or } 4\pi\nu^3/3c^3,$$

so that the total number of independent states of vibration in the frequency range from zero to ν is

$$8\pi\nu^3/3c^3, \quad . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(39·82)}$$

and in the narrow frequency range from ν to $\nu + d\nu$,

$$8\pi\nu^2 d\nu/c^3. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(39·83)}$$

The classical theorem of equipartition of energy (cf. **§ 13·5**) assigns to each vibration an average kinetic energy equal to $kT/2$; or, since in simple harmonic vibrations the average kinetic and average potential energies are equal, the vibrations in the enclosure have the average energy, kT. Combining this with **(39·83)** we find for the energy, $u_\nu\, d\nu$, per unit volume, of the vibrations in the frequency range ν——$\nu + d\nu$

$$u_\nu\, d\nu = 8\pi\nu^2 kT\, d\nu/c^3$$

or

$$u_\nu = 8\pi\nu^2 kT/c^3. \quad . \quad . \quad . \quad . \quad \textbf{(39·84)}$$

This is **Rayleigh's formula.** Indeed, on replacing $u_\nu d\nu$ by $-e d\lambda$, ν by c/λ, and $d\nu$ by $-c d\lambda/\lambda^2$ we get the formula **(39·07).**

§ 39·9. Planck's Quantum Theory

There is no doubt that Rayleigh's formula **(39·07)** is the inevitable outcome of the application of classical dynamical principles to the problem of the spectral distribution of energy in full radiation, and as it does not correctly represent (except in the limiting case of large values of λT) the experimentally determined distribution we are forced, as was Max Planck in 1900, to question their validity. Planck's reflexions on the nature of the modification that the old dynamical principles ought to undergo led him to the view that energy exchanges, at all events when they occur between simple harmonic systems,

occur in integral multiples of $h\nu$, where h is a very small universal constant. It will turn out, in fact, to be identical with the constant, h, of **§ 39·05** (cf. **§ 9·4**). It will therefore be quite in conformity with the spirit of Planck's theory if we deduce an expression for the average energy of a vibration from the assumption that the energy of each vibration of **§ 39·8** is an integral multiple of $h\nu$. We may express this hypothesis of Planck in another way. The Hamiltonian expression for the energy of the simple harmonic vibrations of a particle of mass, m, is

$$H = \frac{p^2}{2m} + \frac{\mu q^2}{2}, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(39·9)}$$

where p $(= mv)$ is the momentum of the particle, q is its displacement at any instant from its equilibrium position, and μ is the restoring force per unit displacement. We may write the equation in the form

$$\frac{p^2}{2mH} + \frac{q^2}{2H/\mu} = 1, \quad . \quad . \quad . \quad . \quad \textbf{(39·901)}$$

so that if we plot p against q we get an ellipse whose semi-axes are

$$(2mH)^{1/2} \text{ and } (2H/\mu)^{1/2}.$$

The area, A, of this ellipse is

$$A = 2\pi H\sqrt{m/\mu},$$

and since the period of oscillation, τ, of the simple harmonic vibration is

$$\tau = 2\pi\sqrt{m/u}$$

we infer that

$$H = A/\tau = A\nu. \quad . \quad . \quad . \quad . \quad \textbf{(39·91)}$$

We can now see clearly the nature of Planck's innovation. It applied specifically to simple harmonic motion and consisted in restricting the constant, A, which classical principles permit to have *any* real and positive value, to integral multiples of a certain, rather small, universal constant, h. It will be observed at once that this makes no practical difference to vibrations on a macroscopic scale and is therefore as it should be, since all the indications point to the adequacy of classical principles when applied to macrophysical phenomena.

Therefore

$$A = nh \quad . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(39·92)}$$

where n is a positive integer or zero and the ellipses **(39·901)** divide up the phase-space into small areas $\iint dp\,dq$ each equal to h. The points in the phase continuum which actually repre-

sent simple harmonic vibrations are points on these ellipses. We may suppose the vibrations for which $n = 0$ to be associated with the innermost ellipse, those for which $n = 1$ to be associated with the area, h, contained between this ellipse and the second one, and so on. The successive areas, each equal to h, correspond to the elements $\Delta\omega_1$, $\Delta\omega_2$, $\Delta\omega_3$, . . . of **§ 13·1**; but whereas these latter elements had no precise value assigned to them and were assumed to approach the limit zero, the elements with which we are now concerned have the precise value, h. There is one further small difference between the elementary areas, h, and the $\Delta\omega_1$, $\Delta\omega_2$, etc., of **§ 13·1**: the points representing the dynamical systems we are now contemplating are situated on the boundaries separating the elementary areas, whereas in **§ 13·1** they were regarded as uniformly distributed through them.

We are now able to apply the results of **§ 13·1** to the calculation of the average energy of the states of vibration with the frequency, ν, in the vacuous enclosure. It is equal to

$$\sum_0^\infty E_s f_s \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(39·93)}$$

where f_s has the meaning defined in that section, $E_s = sh\nu$ and s is a positive integer including 0. Now

$$f_s = Be^{-\frac{sh\nu}{kT}}$$

and

$$B = 1/\sum_0^\infty e^{-\frac{sh\nu}{kT}}$$

or

$$B = 1 - e^{-\frac{h\nu}{kT}}.$$

The average energy required is therefore equal to

$$B\left\{h\nu e^{-\frac{h\nu}{kT}} + 2h\nu e^{-\frac{2h\nu}{kT}} + \ldots\right\}$$

or

$$E = h\nu/\left(e^{\frac{h\nu}{kT}} - 1\right). \quad . \quad . \quad . \quad . \quad \textbf{(39·94)}$$

The only important difference between the method now described and the classical one is that in the latter the areas, $\Delta\omega$, are made indefinitely small and indeed, if we imagine the area h in **(39·94)** to be replaced by one which approaches the limit zero, we get the result kT already used in **§ 39·8.**

To get the new distribution formula we must multiply **(39·83)**, not by kT, but by **(39·94)**. We thus obtain

$$u_\nu = 8\pi h\nu^3/c^3\left(e^{\frac{h\nu}{kT}} - 1\right), \quad . \quad . \quad . \quad \textbf{(39·95)}$$

which is Planck's formula.

If now we write

$$u_\nu \, d\nu = - \, e \, d\lambda$$

and remember that

$$d\nu = - \, c d\lambda / \lambda^2$$

we get

$$e = 8\pi ch/\lambda^5 \left(e^{\frac{ch}{k\lambda T}} - 1 \right)$$

which is identical with **(39·096).**

The introduction and application of the quantum theory to the phenomena of full radiation marks the beginning of a new era in the investigation of physical phenomena just as Einstein's theory of relativity may be said to mark the end of the classical era.

§ 40. Thermal Capacity of Solid Elements

This is related theoretically to full radiation and it is appropriate that the study of it should follow here. The internal energy in a body may be identified with the sum total of the energies of its ultimate particles, and when it is a solid element we may identify these ultimate particles with its constituent atoms and suppose them to be capable of simple harmonic vibrations about their positions of equilibrium. From the classical point of view we must assign to each atom an average kinetic energy, $kT/2$, per degree of freedom in accordance with the principle of equipartition of energy, and therefore $3kT/2$ altogether. In simple harmonic motion the average kinetic and average potential energies are equal, so that the average energy per atom is $3kT$. If now N be the total number of atoms in a gram-atom (Loschmidt's or Avogadro's number, $6{\cdot}0 \times 10^{23}$, approximately) the internal energy in a gram-atom must be

$$3NkT. \quad . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(40)}$$

Now Nk is equal to the gas constant, R, for 1 gramme molecule ($8{\cdot}3 \times 10^7$ ergs per °C. approximately). Consequently the atomic heat at constant volume, according to the classical theory, must be

$$\left.\begin{aligned} 3R &= 24{\cdot}9 \times 10^7 \text{ ergs per °C.} \\ &= \text{roughly 6 cal. per °C.} \end{aligned}\right\} \quad . \quad . \quad \textbf{(40·01)}$$

We have here a very simple explanation of the law of Dulong and Petit. Like Rayleigh's spectral distribution law, it approaches the truth as we approach the macroscopic scale. It is nearer to the results of measurements in the case of elements of high atomic weight and deviates widely from the truth, at ordinary temperatures, in the case of elements of very low atomic weight, carbon, for example. Measurements of the specific

heat of this element (in the form of graphite) have been carried out at relatively high temperatures (as high as 1000° C.) and it has been established that at temperatures of this order its behaviour approximates to the law of Dulong and Petit—another instance of the adequacy of classical physical theory for phenomena on a sufficiently macroscopic scale.

§ 40·1. Debye's Theory

After a tentative effort by Einstein to apply Planck's quantum theory to the problem of the deviations from the law of Dulong and Petit, a high degree of success was achieved by Debye, whose theory of the internal energy of solid elements has an importance and significance which transcends the narrow limits of its more immediate application. It closely resembles the theory of full radiation as given in the foregoing sections. We may regard the material of the solid element as continuous, so long as we are dealing with portions of it which are sufficiently large, and we shall assume—rather tentatively at first—that the internal energy in it is similar to that in the vacuous enclosure of **§ 39·8.** We shall suppose it to be associated with simple harmonic vibrations of an immense range of frequencies extending from very low frequencies to very high ones. These are of course vibrations of the material of the element, not the electromagnetic vibrations of **§ 39·8.** In a unit cube of the material we shall have

$$\frac{8\pi\nu^2 d\nu}{v_\tau{}^3}$$

vibrations, of the transverse kind, whose frequencies lie between ν and $\nu + d\nu$. v_τ means the velocity of transverse waves in the material and is therefore equal to $\sqrt{n/\rho}$, where n and ρ are respectively the rigidity and the density of the material. Similarly there will be

$$\frac{4\pi\nu^2 d\nu}{v_\lambda{}^3}$$

vibrations of the longitudinal kind whose frequencies lie in the range from ν to $\nu + d\nu$. v_λ is the velocity of longitudinal waves in the material and is equal to $\sqrt{(k + 4n/3)/\rho}$, where k is the bulk modulus.[1] The energy in the unit volume will therefore be

$$\int_0^{\nu_m} \left\{\frac{4\pi\nu^2 d\nu}{v_\lambda{}^3} + \frac{8\pi\nu^2 d\nu}{v_\tau{}^3}\right\} \frac{h\nu}{e^{\frac{h\nu}{kT}} - 1} \quad . \quad . \quad . \quad . \quad (40\cdot1)$$

[1] This modulus, k, must not be confused with the entropy constant.

The upper limit, ν_m, may be regarded for the moment as a symbol of our ignorance. If the material were *perfectly* continuous the upper limit in the integral would be infinity, just as in the case of full radiation (cf. **§ 39**).

Of course the integral **(40·1)** cannot include *all* the internal energy; it omits, for example, the contribution due to electromagnetic vibrations. This, however, is of no consequence since the velocity of electromagnetic waves is so large compared with either v_λ or v_τ. Of one thing we may be confident: The choice to make for the upper limit, ν_m, must be such that **(40·1)** is in accord with the classical formula **(40)** at very high temperatures. We shall therefore equate **(40·1)** to $3N'kT$ where N' means the number of atoms in the unit volume and T is made to approach the limit infinity. We thus obtain

$$3N' = \int_0^{\nu m} 4\pi \left\{ \frac{1}{v_\lambda{}^3} + \frac{2}{v_\tau{}^3} \right\} \nu^2 d\nu,$$

or

$$\nu_m{}^3 = 9N'/4\pi \left\{ \frac{1}{v_\lambda{}^3} + \frac{2}{v_\tau{}^3} \right\}. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(40·11)}$$

The frequency ν_m is therefore determined by the elastic constants of the material, the number of atoms in the unit volume and the density, and may now be assumed to be known. If we make use of **(40·11)** we can express the internal energy of the unit volume **(40·1)** more compactly as follows:

$$\frac{9N'}{\nu_m{}^3} \int_0^{\nu_m} \frac{h\nu^3 d\nu}{e^{\frac{h\nu}{kT}} - 1} \quad . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(40·12)}$$

and we get the appropriate expression for the internal energy of a gram-atom by replacing N' by $N(= 6{\cdot}0 \times 10^{23})$ the number of atoms in a gram-atom. As in **§ 39·5** and the following sections, we shall use the symbol, β, as an abbreviation for $h\nu/kT$ and β_m for $h\nu_m/kT$ so that the internal energy of a gram-atom may be written

$$\frac{9Nk^4T^4}{h^3\nu_m{}^3} \int_0^{\beta m} \frac{\beta^3 d\beta}{e^\beta - 1}. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(40·13)}$$

Whatever value ν_m may happen to have for a particular element, β_m must approach the limit ∞ as T approaches the limit zero, so that at very low temperatures the integral in **(40·13)**

approaches the numerical value $6 \times 1{\cdot}0823$ (cf. **§ 39·6**) and the **atomic heat** becomes:

$$4 \times 9N \times 6 \times 1{\cdot}0823k^4T^3/h^3\nu_m^3. \quad . \quad . \quad \textbf{(40·14)}$$

It is proportional to the cube of the absolute temperature. Between the two extremes, **(40·14)** and **(40·01)**, the atomic heat depends on the temperature in a way which can best be described in the following manner: We write **(40·13)** conveniently in the form:

$$9NkTf(\beta_m), \quad . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(40·15)}$$

where $$f(\beta_m) \equiv \frac{1}{\beta_m^3}\int_0^{\beta_m} \beta^3 d\beta/(e^\beta - 1).$$

The atomic heat is therefore

$$9Nkf(\beta_m) + 9NkTf'(\beta_m)d\beta_m/dT,$$

where $f'(\beta_m)$ means $df(\beta_m)/d\beta_m$.

Hence the atomic heat may be expressed as

$$9Nk\{f(\beta_m) - \beta_m f'(\beta_m)\},$$

since $d\beta_m/dT = -\beta_m/T$, or more shortly

$$9Nk\phi(\beta_m), \quad . \quad . \quad . \quad . \quad . \quad \textbf{(40·16)}$$

where $\phi(\beta_m)$ is an abbreviation for $f(\beta_m) - \beta_m f'(\beta_m)$.

Now ϕ is the same function for all elements and

$$1/\beta_m = T/(h\nu_m/k)$$

or

$$1/\beta_m = T/\Theta \quad . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(40·17)}$$

where $\Theta \equiv h\nu_m/k$. So that we may describe $1/\beta_m$ as the temperature measured by the unit, Θ, which is called the **characteristic temperature** of the element. **It appears then that when the atomic heat is plotted against $1/\beta_m$ the same curve represents the dependence of the atomic heat on the temperature for all elements.** A word of caution must be said here. The atomic heat, as we have defined it, is measured by the increase in internal energy per unit rise in temperature, **the latter being expressed in terms of the same unit (the Centigrade degree on Kelvin's work scale) for all elements;** but in the curve of Fig. 40·1 the

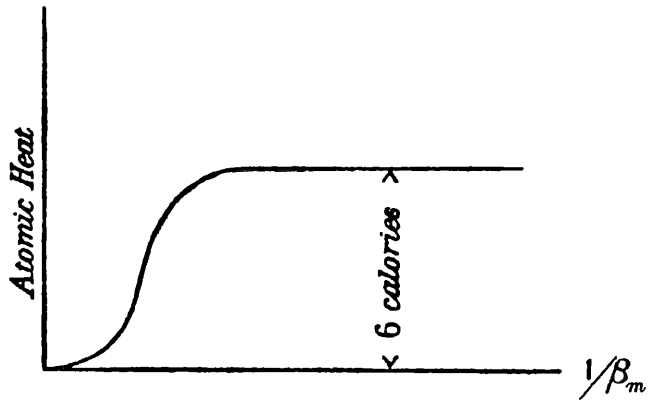

FIG. 40·1

atomic heat, so measured, is plotted against the temperature as measured by $1/\beta_m$.

One feature of Debye's theory is especially worthy of notice: **the atomic heat of a solid element can be calculated from the values of its elastic constants.** This is due of course to the fact that ν_m is determined in each case (cf. **40·11**) by the elastic constants k and n, since they determine the values of v_λ and v_τ.

Debye's theory has been found to be very accurately in accord with the results of experiments on copper, silver, zinc, and other monatomic elements, and the portionality with T^3 at low temperatures is well supported. It has been slightly extended by Born and Karman, who have taken into account that solid substances are crystalline and have therefore characteristic modes of vibration determined by their crystalline structure.

It will have been noticed how closely Debye's theory is related to that of full radiation. The two are indeed identical and we may say that this theory of the thermal capacity of solid elements is almost, if not quite, as much a tribute to the genius of Planck as to that of Debye himself. Its success is due, however, not only to the fact that it is based on Planck's foundations, but also to the assumption, which happens to be true over a wide range of temperatures, that the heat energy of solids can be identified with the energy of their elastic vibrations. At extremely low temperatures, however—temperatures of a fraction of a degree on Kelvin's work scale—this energy of vibration becomes negligible, or at all events ceases to vary with the temperature. At these very low temperatures the energy of the electronic systems of the atoms and more particularly that of their magnetic fields becomes dominant.

§ 40·2. Thermal Capacity of Gases

We have studied the thermal capacities of two media, namely, of free space and of a solid element. Though these two cases are in many respects very different from one another, they have this in common: we may regard the internal energy under temperature equilibrium as distributed among simple harmonic vibrations whose frequencies range from zero to very high values —infinity in the case of free space. Such discrepancies as we find between the pronouncements of classical theory and those of experimental observations are more and more marked the lower the temperature. The interesting result **(39·14)**, known as the Stefan-Boltzmann law, was obtained without departing from classical methods. This was possible because of the fact

that the internal energy density is necessarily a function of the temperature only and the further fact that the pressure is known to be equal to one-third of the energy density. The laws of thermodynamics could therefore be brought into operation without the aid of constitutive hypotheses. The corresponding problem for the solid element, as dealt with by classical methods, is only possible with the aid of a constitutive hypothesis: namely, that of atoms capable of simple harmonic motions about fixed positions of equilibrium. It is this hypothesis which leads to the law of Dulong and Petit, a law which fails more and more conspicuously the lower the temperature.

We may illustrate this kind of behaviour at low temperatures by studying the case of a gas. The constitutive hypothesis is now that of molecules, which are almost free from forces except such as are evoked in collisions, and associated with it we have again, when we use the classical theory, the principle of equipartition of energy which requires that the average kinetic energy of each molecule is $kT/2$ per degree of freedom, k being the familiar constant we have met in **§§ 12, 12·1, 13·5** and elsewhere. The specific heat of the gas at constant volume is therefore approximately (according to the classical kinetic theory, **§ 12**) equal to

$$Nnk/2,$$

where N is the number of molecules per unit mass and n is the number of degrees of freedom of the molecule, or

$$nR/2,$$

where R is the gas constant per unit mass. In this expression the heat is of course expressed in work units. There is the further consequence which may be noted here, namely,

$$c_p/c_v = 1 + \frac{2}{n}$$

(cf. **12·03**).

Experiment indicates that, as the temperature is lowered, the ratio (c_p/c_v) approaches the limit $1\frac{2}{3}$ which, according to the classical kinetic theory, is the value appropriate to a monatomic gas. We see, therefore, that the lowering of the temperature has the same effect on the gas as if all its degrees of freedom, except those associated with translatory motion, were to disappear. The older quantum theory explained this in a fairly simple way, which we shall now describe.

Planck's original quantum hypothesis assigned the energy $nh\nu$ to a simple harmonic motion of one degree of freedom, n being an integer. An alternative way of expressing this hypothesis is indicated in **§ 39·9.** The relation—in simple harmonic motion

—between the momentum, **p**, and the positional co-ordinate, **q**, is expressed by **(39·901)**. The area of the ellipse, which the equation describes must be equal, when Planck's expression, **nhν**, for the energy is adopted, to nh. Consequently

$$\oint p dq = nh, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(40·2)}$$

is an alternative expression of Planck's hypothesis. In this form it is capable of extension to a large class of dynamical systems—those for which the independent positional co-ordinates q_α can be so chosen that each q_α is associated with a period and the conjugate p_a is a function of the corresponding q_α. The integral $\oint p_\alpha \, dq_\alpha$ is then a constant and **(40·2)** may be extended to

$$\oint p_\alpha \, dq_\alpha = n_\alpha h. \quad . \quad . \quad . \quad . \quad \textbf{(40·201)}$$

There is one of these equations for each q_α. This extension of Planck's hypothesis was suggested in 1915 by the author of this book and was also discovered independently by Arnold Sommerfeld. We shall apply it to investigate the thermal capacity of a gas having 5 degrees of freedom, 3 of which we associate with the possibility of translation and 2 with that of rotation. We must then apply **(40·201)** to each degree of freedom. Beginning with the component of the motion of a molecule parallel to the X axis we have

$$\oint p_\alpha \, dx = n_1 h.$$

Since the momentum of a molecule is practically constant between two consecutive collisions, we may regard its motion as a simple harmonic one with an infinitely long period or infinitely small frequency. Therefore, since $n_1 h\nu$ is not infinitely small, n_1 is always enormously large. Planck's hypothesis therefore does not depart from the classical theory when applied to the translatory motion of the molecules. It is very different when we apply **(40·201)** to the rotational motion of molecules. A rigid body in rotation about a fixed axis has the energy

$$\tfrac{1}{2} I \dot{\theta}^2, \quad . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(40·21)}$$

where $\dot{\theta}$ is its angular velocity about the axis of rotation and I is the corresponding moment of inertia. The condition **(40·201)** requires that

$$\oint I \dot{\theta} \, d\theta = sh,$$

where s is a positive integer or zero. Therefore

$$I\dot{\theta} = \frac{sh}{2\pi},$$

since $I\dot{\theta}$, the angular momentum, is constant. Consequently the energy associated with this rotation is, by **(40·21)**,

$$E_s = \frac{s^2h^2}{8\pi^2 I}. \quad . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(40·22)}$$

Now the average energy of rotation (per degree of freedom) will be (cf. **13·101**)

$$\bar{E} = \Sigma E_s f_s, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(40·23)}$$

where $$f_s = e^{-E_s/kT} / \sum_s e^{-E_s/kT}$$

or

$$f_s = e^{-\sigma s^2} / \sum_s e^{-\sigma s^2}, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(40·24)}$$

where

$$\sigma \equiv h^2/8\pi^2 IkT. \quad . \quad . \quad . \quad . \quad \textbf{(40·241)}$$

Therefore the average kinetic energy of rotation per degree of freedom is, **(40·23)**,

$$\bar{E} = \frac{\sum \frac{s^2h^2}{8\pi^2 I} e^{-\sigma s^2}}{\Sigma e^{-\sigma s^2}}$$

or

$$\bar{E} = \sigma kT \frac{\Sigma s^2 e^{-\sigma s^2}}{\Sigma e^{-\sigma s^2}}, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(40·25)}$$

which may obviously be written

$$\bar{E} = -\sigma kT \frac{d}{d\sigma} \log \sum_s e^{-\sigma s^2}. \quad . \quad . \quad . \quad \textbf{(40·251)}$$

This expression is in accord with the principle of equipartition under macrophysical conditions, i.e. when we may regard h as approaching the limit zero, or also when T is very great. Under these circumstances σ approaches the limit zero, and consequently, if we define α by

$$\alpha \equiv \sigma^{1/2} s,$$

we shall have

$$\sigma^{1/2} = d\alpha$$

and therefore

$$\sum_1^\infty{}_s e^{-\sigma s^2} = \sigma^{-1/2} \int_0^\infty e^{-\alpha^2}\, d\alpha$$

$$= \tfrac{1}{2}\sigma^{-1/2}\pi^{1/2}. \quad . \quad . \quad . \quad . \quad \textbf{(40·252)}$$

Therefore in the macrophysical case

$$\bar{E} = -\sigma kT\frac{d}{d\sigma}\log\left(\tfrac{1}{2}\sigma^{-1/2}\pi^{1/2}\right)$$

or finally

$$\bar{E} = kT/2$$

as we should expect.

The formula **(40·251)** may be written

$$\bar{E} = -\frac{h^2}{8\pi^2 I}\frac{d}{d\sigma}\log\sum_s e^{-\sigma s^2}$$

and when we apply it to a gas with, for example, two rotational degrees of freedom, we find for its internal energy

$$\tfrac{3}{2}RT - 2.\frac{h^2}{8\pi^2 I}.N.\frac{d}{d\sigma}\log\Sigma,$$

where R is the gas constant, N is the number of molecules, and Σ is the sum $\sum_1^\infty e^{-\sigma s^2}$. And for the specific heat at constant volume we get

$$C_v = \tfrac{3}{2}R + 2R\sigma^2\frac{d^2\log\Sigma}{d\sigma^2}, \quad . \quad . \quad . \quad \textbf{(40·26)}$$

the constant, R, now being that appropriate to the unit mass of the gas.

As the temperature approaches the limit zero, and consequently σ approaches infinity, the second term in **(40·26)** approaches zero.

Thus the old-fashioned quantum theory provides an explanation of the observational fact that the specific heat (at constant volume) of a gas diminishes with diminishing temperature in such a way as to suggest that the number of degrees of freedom per molecule tends to be reduced to 3, the number associated with the molecules of monatomic gases.

CHAPTER IX

DEVELOPMENT OF THE QUANTUM THEORY TILL THE EMERGENCE OF WAVE MECHANICS

§ 40·3. Photo-electricity

THE further development of the quantum theory following the achievements of Planck in the investigation of the distribution problem of full radiation and of Debye in the problem of the thermal capacity of solid elements was linked up with the study of photo-electricity and of the regularities found in spectra, more especially the series spectra of hydrogen, helium and the alkali metals, lithium, sodium, etc.

The outstanding fact about the photo-electrical phenomenon, i.e. the phenomenon of the emission of electricity, in the form of electrons, under the influence of radiation (light or X-radiation) is that the maximum kinetic energy of the individual electrons ejected is independent of the intensity of the illumination of the emitting metallic or other surface. The number of electrons ejected is proportional to the radiant energy reaching the emitting surface; but the violence with which they are thrown out is quite independent of this and is just as great under the influence of the weakest illumination as under that of very intense illumination. It is characteristic of the photo-electric effect that the energy of the ejected electrons is determined by the frequency, ν, of the exciting radiation and that there is for each material a fairly definite **threshold frequency** below which no emission of electrons occurs. Einstein explained these puzzling features of the phenomenon by supposing light to have a corpuscular constitution (cf. **§§ 34·6, 39**), each corpuscle or photon having the energy $h\nu$. This hypothesis is quite effective in explaining the observed facts of photo-electric phenomena. If we suppose that when an electron is ejected all the energy of a photon is used up, we get the equation

$$h\nu = \tfrac{1}{2}mv^2 + \phi, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(40·3)}$$

where ϕ is the work that has to be done on the electron to drag it out of the illuminated material. There will be a certain frequency, ν_0, characteristic of the material, for which

$$h\nu_0 = \phi$$

and for which therefore the kinetic energy of the emerging electron vanishes. Photons of lower frequency than this will not be able to deliver sufficient energy to an electron to enable it to emerge through the illuminated surface. Thus ν_0, is the threshold frequency.

In the case of X or gamma radiation, the photons of which have large energies (ν is very large compared with ν_0), equation **(40·3)** becomes practically identical with

$$\left.\begin{aligned} h\nu &= \tfrac{1}{2}mv^2, \\ \text{or} \qquad h\nu &= m_0c^2(\gamma - 1). \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad \textbf{(40·31)}$$

All these consequences of Einstein's hypothesis are entirely supported by experimental results; but it raised a formidable difficulty in connexion with interference and diffraction phenomena, which are observable even with light the intensity of which is so small that any possibility of accounting for them by overlapping photons is quite excluded—even if such an explanation were permissible, implying as it does that the photons are small coherent wave-groups. The solution of this problem is accomplished by the wave-mechanics of L. de Broglie and E. Schroedinger which we shall study in a later chapter (**§§ 41·6** et seq.).

The excitation of X-rays by electron bombardment is the converse phenomenon to the photo-electric one. The energy of the colliding electron is practically equal to eV, where e is the numerical value of the electronic charge and V is the potential difference between the cathode and the anti-cathode. We have therefore for the *upper limit* of the frequency, ν, of the emitted X-radiation—since $h\nu$ is the energy of an X-ray photon:

$$h\nu = eV. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(40·32)}$$

X-radiation of lower frequency may be, and is, excited, since the whole of the energy eV is not always used in the ejection of a photon. It may be used partly too in changing the electronic state or constitution of the atoms in such a way that they become capable of emitting radiation of lower frequency than the upper limit given by **(40·32)** (Barkla's **characteristic radiation**).

§ 40·4. The Compton Effect

The suggestion of a *quasi* corpuscular constitution for electromagnetic radiation is strongly supported by the phenomenon known as the **Compton effect.** Indeed, its discoverer, A. H. Compton, was led to search for it because it appeared to be a consequence of the view that light (and X-radiation in par-

ticular) has a corpuscular structure, each corpuscle (photon) having the energy $h\nu$. Though now known as the *Compton* effect it appears to have been discovered experimentally by C. G. Barkla many years earlier. He noticed that the X-radiation scattered by various materials was invariably slightly more absorbable in aluminium than the primary radiation, a fact which reveals its slightly longer wave-length.

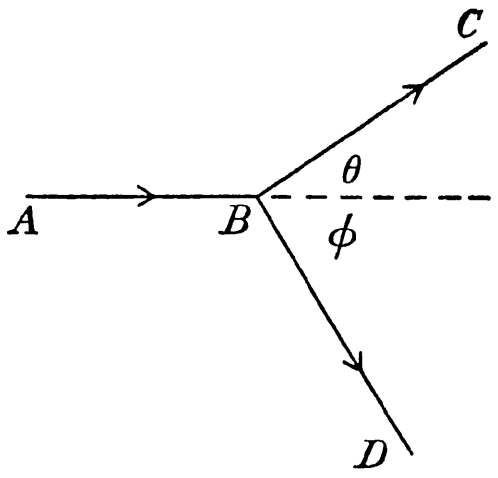

FIG. 40·4

Imagine a photon to collide with an electron, the latter being assumed to be at rest before the collision, and let the line along which the photon is travelling before colliding with the electron be AB, the electron being, before the collision, at rest at B. If after the collision the scattered photon travels away along some line BC, while the electron travels along BD, these lines making the angles θ and ϕ respectively with AB, we easily derive the following equations. From the energy principle

$$h\nu = h\nu' + m_0c^2(\gamma - 1), \quad . \quad . \quad . \quad . \quad \textbf{(40·4)}$$

while the principle of momentum yields

$$\left.\begin{aligned} \frac{h\nu}{c} &= \frac{h\nu'}{c}\cos\theta + m_0\gamma v\cos\phi, \\ 0 &= \frac{h\nu'}{c}\sin\theta - m_0\gamma v\sin\phi. \end{aligned}\right\} \quad . \quad . \quad \textbf{(40·41)}$$

In these equations ν is the frequency of the incident photon, ν' is its frequency after its collision with the electron, m_0 is the rest mass of the electron, and γ is the familiar factor $(1 - v^2/c^2)^{-1/2}$ of relativity theory. The expressions $h\nu/c$ and $h\nu'/c$ for the momentum before and after collision naturally suggest themselves, since

$$mc^2 = h\nu$$

and hence

$$mc = h\nu/c,$$

m being the mass of the photon. The equations **(40·41)** have been derived with the tacit assumption that AB, BC, and BD lie in the same plane. A very little reflexion will make it evident that the principle of conservation of momentum requires this.

From **(40·4)** we easily get

$$m_0\gamma = \frac{h(\nu - \nu')}{c^2} + m_0$$

and consequently

$$m_0{}^2\gamma^2 = \frac{h^2}{c^4}(\nu - \nu')^2 + \frac{2m_0h}{c^2}(\nu - \nu') + m_0{}^2, \quad . \quad \textbf{(40·42)}$$

while from equations **(40·41)** we obtain

$$\frac{h^2}{c^2}(\nu^2 + \nu'^2 - 2\nu\nu' \cos\theta) = m_0{}^2\gamma^2v^2,$$

or since $\gamma^2v^2 = c^2(\gamma^2 - 1)$,

$$m_0{}^2\gamma^2 = \frac{h^2}{c^4}\left\{(\nu - \nu')^2 + 4\nu\nu' \sin^2\frac{\theta}{2}\right\} + m_0{}^2. \quad . \quad \textbf{(40·43)}$$

On combining this with **(40·42)** we get

$$\nu - \nu' = \frac{2h}{m_0c^2}\nu\nu' \sin^2\frac{\theta}{2},$$

and since $\nu = c/\lambda$ and $\nu' = c/\lambda'$, where λ and λ' are respectively the wave-lengths associated with the incident and the scattered photons,

$$\boxed{\lambda' - \lambda = \frac{2h}{m_0c} \sin^2\frac{\theta}{2}.} \quad . \quad . \quad . \quad . \quad \textbf{(40·44)}$$

Compton measured the change in wave-length, $\lambda' - \lambda$, due to scattering, of X-rays for various values of the angle, θ, and established the substantial correctness of this equation **(40·44)**.

The assumptions just made for the energy, $h\nu$, and the momentum, $h\nu/c$, of a photon are just those suggested by the analogy between classical dynamics and geometrical optics (**§§ 9·4** and **41·5**) which leads to h/τ (τ is the period) for the energy and h/λ for the momentum of any particle.

Without any calculation we may infer that the radiation scattered by electrons must have a longer wave-length after scattering, since the photons must impart some of their energy to the electrons—at all events when the latter are initially practically at rest. Hence the product $h \times$ *frequency* must be smaller for the scattered than for the incident photons. Their frequency is therefore smaller and their wave-length consequently longer.

§ 40·5. Elementary Facts about Series Spectra

As early as the year 1885 Balmer discovered that the wave-lengths of the lines in the most prominent series in the spectrum

of the light from glowing hydrogen could be represented with accuracy by the formula

$$\lambda = A\frac{n^2}{n^2 - 4}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (40{\cdot}5)$$

in which λ is the wave-length of a line, A is a constant to which Balmer gave the value 3645·6 in Angstrom units (one A.U. is 10^{-10} metres), and n is any integer in the sequence 3, 4, 5, . . . ∞. Using the value of A just given the wave-lengths corresponding to $n = 3$, $n = 4$, $n = 5$, and $n = 6$—they are usually called the α, β, γ, and δ lines—are

$$\begin{aligned} \lambda_\alpha &= 6562{\cdot}08 \text{ A.U.} \\ \lambda_\beta &= 4860{\cdot}80 \quad ,, \\ \lambda_\gamma &= 4340{\cdot}00 \quad ,, \\ \lambda_\delta &= 4101{\cdot}30 \quad ,, \end{aligned}$$

It will be noticed that the length, A, is the limit which the wave-lengths of the light corresponding to the various lines in the series approach as n approaches infinity.

Balmer's success with hydrogen was followed by the discovery of other important regularities in spectra, notably by Rydberg and Ritz. The former gave Balmer's law the form:

$$\nu' = \frac{4}{A}\left(\frac{1}{4} - \frac{1}{n^2}\right),$$

or

$$\nu' = R'\left(\frac{1}{2^2} - \frac{1}{n^2}\right), \quad . \quad . \quad . \quad . \quad (40{\cdot}501)$$

where ν' is the **wave-number**—i.e. the number of waves in the unit length. The constant R', known as **Rydberg's constant,** is equal (if we adopt Balmer's value of A) to

$$4/3645{\cdot}6 \text{ (A.U.)}^{-1}$$

or

$$R' = 109721 \text{ cm.}^{-1} \quad . \quad . \quad . \quad . \quad (40{\cdot}502)$$

approximately. The calculated wave-lengths, in vacuo, of the lines in Balmer's series fit the observations best when we adopt for R' the value 109677·6 cm.$^{-1}$.

It is sometimes convenient to multiply both sides of **(40·501)** by c (= approximately 3×10^{10} cm./sec.) so that we have

$$\nu = R\left(\frac{1}{2^2} - \frac{1}{n^2}\right), \quad . \quad . \quad . \quad . \quad (40{\cdot}503)$$

in which ν is the **frequency** or the number of waves emitted in the unit time and R (which is also called Rydberg's constant) is equal to **cR'**. Obviously

$$R = 3{\cdot}29 \times 10^{15} \text{ sec.}^{-1} \quad . \quad . \quad . \quad (40{\cdot}504)$$

approximately.

A very slight generalization of **(40·503)** gives us

$$\nu = R\left(\frac{1}{n_0^2} - \frac{1}{n^2}\right) \quad . \quad . \quad . \quad . \quad . \quad \textbf{(40·51)}$$

and when we adopt the values 1, 2, 3, and 4 respectively for n_0 the resulting formulae do in fact give the wave-lengths for the lines of four series which have been observed in the hydrogen spectrum, namely Lyman's series, Balmer's series, Paschen's series, and Brackett's series.

The form of **(40·51)** suggested to Ritz one of the most important and fundamental spectral laws: namely, that the frequencies (and therefore also the wave-numbers) of the light emitted by a system (atom or molecule) can be expressed as the differences of numerical quantities, called **spectral terms,** which are characteristic of the particular emitting atom or molecule. In the case of the hydrogen atom these terms are represented by R/n^2, where R is Rydberg's constant and n is any positive integer from 1 to ∞. This important generalization is known as the **combination principle** of Ritz. While it is true that every observed frequency (or wave-number) is equal to the difference of two of the spectral terms which belong to the emitting atom or molecule the converse statement is not true. That is to say, if we take any pair of the spectral terms of an atom their difference is not always equal to the frequency of a spectral line of the atom.

Balmer's series is typical of the various series of lines which are observed in the emission spectra of atoms. Such series are represented by a formula containing a constant term, $R/2^2$, in the special case of Balmer's series, and a variable one, R/n^2, in the series spectra of hydrogen, and one term is distinguished from another by the value of some integer (or it may be integers), the integer n in the case of Balmer's series. At an early period in the history of spectroscopic investigation some of these series were called **principal series,** others **sharp series, diffuse series,** and so on, on account of their prominence or the appearance of the lines in them. Consequently the variable terms in the formulae representing the wave-numbers (or frequencies) of the lines in these series were called P terms, S terms, D terms, and so on. A further spectroscopic discovery was that the letters representing these different classes of spectral terms can be placed in such an order: for example

$$S,\ P,\ D,\ F,\ \ldots,$$

that only those differences of spectral terms represent observed frequencies when the terms belong to classes denoted by neigh-

bouring letters. Thus the difference between an S term and a D term never represents a spectral frequency, but the difference between an S term and a P term does, or that between a P term and a D term. It is convenient for our immediate purely descriptive purpose to represent the spectral terms of an atom, e.g. sodium, collectively in the following way

$$\begin{array}{l} 1S,\ 2S,\ 3S,\ \ldots \\ 1P,\ 2P,\ 3P,\ \ldots \\ 1D,\ 2D,\ 3D,\ \ldots \\ 1F,\ 2F,\ 3F,\ \ldots \end{array}$$

If we do this the frequencies (or wave-numbers) of the principal series of sodium are represented by the sequence

$$1S - 1P,\ 1S - 2P,\ 1S - 3P,\ \ldots$$

the upper limit of the sequence being $1S$. Indeed, the term nP may be represented, quite roughly, by $R/(n + \mu)^2$ where R is Rydberg's constant, n is the serial integer, and μ, though not quite a constant, varies very little. The sequence resembles, as is the case with all the alkali metals, that of Balmer's series. With this provisional notation the sharp series of sodium would be represented by

$$1P - 2S,\ 1P - 3S,\ 1P - 4S,\ \ldots$$

and the diffuse series by

$$1P - 1D,\ 1P - 2D,\ 1P - 3D,\ \ldots$$

Each of the two latter series has the common upper limit $1P$, and this limit subtracted from the upper limit $1S$ of the principal series gives $1S - 1P$ the first line of the principal series:

[Limit of Principal Series]
− [Limit of Sharp or Diffuse Series]
= [First Line of Principal Series].

This is known as the **Rydberg-Schuster law.**

In the present notation

$$nS > nP > nD > nF\ \ldots,$$

n being the same (serial) integer in each case.

There is one important feature of these series which has been tacitly ignored till now: every line in each series is multiple. In the case of the alkali metals it is double (e.g. the well-known D lines of sodium which constitute the first doublet member of its principal series). The separation of the members of each doublet (as measured by their wave-numbers) diminishes progressively in a principal series as the serial integer increases and approaches the limit zero as n approaches ∞. In the case of

the sharp series the separation is constant. These peculiarities were of course discovered long before Ritz's combination principle. They are due to the fact that the S terms are all single while each P, D, etc., term is double—in the case of the alkali metals. That is to say, we find not one but two P terms differing little one from the other. Since the P terms diminish with the serial integer n the diminution of the frequency difference between the members of a doublet in a principal series is a necessary consequence. Similarly, since the doublets in a sharp series are represented, each one, by subtracting a single S term from a constant double P term, the doublet separation is constant through the series.

§ 40·6. Bohr's Derivation of Rydberg's Constant

In the year 1913 Niels Bohr succeeded in deducing the formula **(40·51)** and in calculating the value of Rydberg's constant from premisses provided by Planck's quantum theory. He hit upon the brilliant idea of associating the energy values of the emitting system (atom or molecule) with its spectral terms. His chief postulate may be expressed in the form:

$$\textbf{spectral term} = \textbf{energy}/h, \quad . \quad . \quad . \quad \textbf{(40·6)}$$

when the frequencies, ν, are being calculated, or

$$\textbf{spectral term} = \textbf{energy}/ch, \quad . \quad . \quad \textbf{(40·601)}$$

when wave-numbers, ν', are being calculated. The constants c and h are of course the velocity of light in vacuo and Planck's constant respectively. Thus the frequency of a line is the difference of two spectral terms **(40·6)** and is therefore equal to the difference of the energies of two stationary states divided by h. In other words, the energy difference, which represents the energy emitted (or absorbed), is equal to $h\nu$. This is completely in accord with Planck's principles and also with the view now held that light consists of *quasi* particles each having the energy $h\nu$. Bohr adopted Rutherford's view that all atoms consist of an inner, relatively massive, positively charged nucleus with planetary electrons revolving round it, and he assigned to hydrogen the simplest structure of this type, namely a nucleus with a positive charge equal to the elementary charge e ($4{\cdot}774 \times 10^{-10}$ O.E.S.U.) and a single electron with a numerically equal charge. In his earliest work he simplified his calculations by assuming the centre of mass of the atom to be in the nucleus itself—this is permissible as a first approximation, since the mass of a hydrogen atom is about 1845 times as great

as that of an electron—and he confined his attention to circular orbits only.

In describing this first effort to give a theoretical basis for the origin of spectra it is convenient, since we are concerned with energy *differences* only, to fix the constant part, whether we think of it as the m_0c^2 of relativity theory or the arbitrary constant of Hamiltonian mechanics, in such a way that the potential energy of the atomic system is zero when the radius, r, of the electronic orbit is infinite. With this convention the energy, E, of the atom is negative and equal to $-e^2/2r$, while its kinetic energy is equal to $e^2/2r$. We are using e to represent the numerical value of the electronic charge. That is to say e means $+ 4{\cdot}774 \times 10^{-10}$ when measured by ordinary electrostatic units. Bohr adopted the hypothesis

$$E = -nh\omega/2, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(40·61)}$$

where n is a positive integer, h is Planck's constant, ω is the frequency of revolution of the electron in its orbit, i.e. the number of revolutions it makes per second, and E is the energy of the system in one of its stationary states. It will be remembered that Planck's quantum theory regards any system, *while in a stationary state*, as conservative and conforming to Hamiltonian dynamical laws. Therefore

$$mv^2 = 2T = -2E = e^2/r$$

where m is the mass of the electron, v its constant velocity in its circular orbit, T its kinetic energy, and r is the orbital radius. When we combine these equations with **(40·61)** and remember that $\omega = v/2\pi r$ we easily get

$$E = -\frac{2\pi^2me^4}{n^2h^2}, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(40·62)}$$

so that the spectral terms become in virtue of the postulate **(40·6)** (if we agree to make them positive)

$$\frac{2\pi^2me^4}{h^3} \times \frac{1}{n^2}.$$

Thus, according to Bohr

$$R = 2\pi^2me^4/h^3. \quad . \quad . \quad . \quad . \quad \textbf{(40·63)}$$

and when we substitute in this formula the value $\cdot 9 \times 10^{-27}$ for m, $6{\cdot}54 \times 10^{-27}$ for h, and $4{\cdot}77 \times 10^{-10}$ for e the agreement with **(40·504)** is very striking.

The condition **(40·2)** unites the apparently discrepant conditions: $E = nh\nu$ of Planck, $E = -nh\omega/2$ of Bohr and the suggestion, made by Nicholson several years before Bohr suc-

ceeded with hydrogen, that an angular momentum is always equal to $nh/2\pi$.

Let us now deal again with the problem we have just solved — this time without making the approximate assumption that the centre of mass is in the nucleus. For simplicity we shall continue to deal with the special case of circular orbits. The planetary electron travels in a circle of radius r (Fig. 40·6). The nucleus travels in a circle of smaller radius, R, and the centre of mass of the system is at O. Obviously

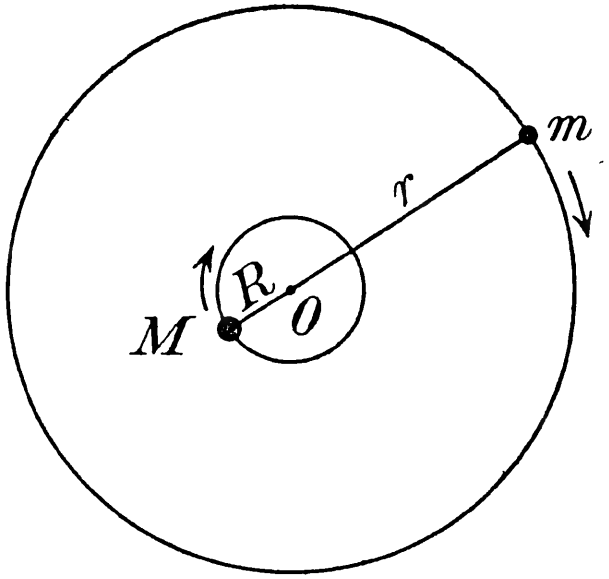

FIG. 40·6

$$\frac{R}{r} = \frac{m}{M} = \frac{V}{v} = s, \quad . \quad . \quad . \quad . \quad \textbf{(40·64)}$$

where m and M are the masses of the electron and of the nucleus respectively, v and V their respective velocities and s is a small constant. The quantum condition of **(40·2)** when applied to this case gives

$$(mvr + MVR)2\pi = nh,$$

or by **(40·64)**

$$mvr(1 + s)2\pi = nh$$

or finally

$$2\pi mvr = \varepsilon nh \quad . \quad . \quad . \quad . \quad . \quad \textbf{(40·65)}$$

where ε means $1/(1 + s)$ and is not very different from unity. We shall make our problem a little more general by supposing the charge on the nucleus to be Ze numerically where Z is a positive integer called the **atomic number.**[1] The centrifugal force on the electron and on the nucleus is balanced by the electric attraction between them. Therefore

$$\frac{mv^2}{r} = \frac{MV^2}{R} = \frac{Ze^2}{(r + R)^2}$$

and therefore

$$mv^2 + MV^2 = \frac{Ze^2}{(r + R)},$$

[1] The term has acquired this meaning in recent times. As originally used it meant the number of the element in the list of elements arranged in the order of increasing atomic weights.

so that

$$-2E = \frac{\varepsilon Z e^2}{r}. \quad . \quad . \quad . \quad . \quad . \quad (40{\cdot}66)$$

Finally, we may note that

$$v = \sqrt{\left|-\frac{2E\varepsilon}{m}\right|} \quad . \quad . \quad . \quad . \quad (40{\cdot}67)$$

and on eliminating v and r from the three equations **40·65, 40·66,** and **40·67** we find

$$E = -\frac{2\pi^2 m\varepsilon Z^2 e^4}{n^2 h^2}. \quad . \quad . \quad . \quad . \quad (40{\cdot}68)$$

This differs from **(40·62)** only in the factor ε (Z is unity in **40·62**) and

$$\varepsilon = 1/(1+s) = M/(M+m) \quad . \quad . \quad (40{\cdot}681)$$

and is always near unity. The spectral terms are now

$$RZ^2/n^2$$

and

$$R = 2\pi^2 m\varepsilon e^4/h^3, \quad . \quad . \quad . \quad . \quad (40{\cdot}69)$$

so that Rydberg's constant varies slightly in value from one element to another.

§ 40·7. Bohr's Calculation of the Ratio of the Masses of the Proton and Electron

The case of singly ionized helium is of rather special interest. The neutral helium atom has a mass which is 3·97 times that of the hydrogen atom. So that we may assume, with a sufficient approach to accuracy, that its nucleus has a mass which is 3·97 times that of the hydrogen nucleus—which is called a **proton.** The positive charge on the helium nucleus is $2e(Z=2)$. The neutral atom has two planetary electrons and the singly ionized atom only one. Bohr's theory suggests for the latter the spectral terms:

$$4R_{\mathrm{He}}/n^2.$$

The subscript He is meant to indicate that we have here the Rydberg constant appropriate to helium. We may anticipate therefore the spectral series included in the general formula

$$\nu = 4R_{\mathrm{He}}\left\{\frac{1}{n_0^2} - \frac{1}{n^2}\right\}. \quad . \quad . \quad . \quad . \quad (40{\cdot}7)$$

If, for example, $n_0 = 4$ we have the series

$$\nu = 4R_{\mathrm{He}}\left\{\frac{1}{4^2} - \frac{1}{n^2}\right\}$$

or

$$\nu = R_{\mathrm{He}}\left\{\frac{1}{2^2} - \frac{1}{(n/2)^2}\right\} \quad . \quad . \quad . \quad . \quad \textbf{(40·71)}$$

in which n is any positive integer for which $n/2$ is greater than two, i.e. $n = 5, 6, 7, 8, 9, \ldots$ For even values of n, therefore, **(40·71)** is practically coincident with Balmer's series of hydrogen. The slight difference between them is due solely to the slight difference between R_{Helium} and R_{Hydrogen}.

The lines corresponding to odd values of n, namely 5, 7, 9 . . ., were observed by Pickering in the spectrum of the star ζ *Puppis* in 1896 and were quite naturally, but erroneously, thought to be due to hydrogen. The slight difference in the values of the R's of helium and hydrogen is of great interest. Evidently

$$\frac{R_{\mathrm{Helium}}}{R_{\mathrm{Hydrogen}}} = \frac{\varepsilon_{\mathrm{Helium}}}{\varepsilon_{\mathrm{Hydrogen}}}$$

or approximately

$$\frac{R_{\mathrm{He}} - R_{\mathrm{H}}}{R} = \varepsilon_{\mathrm{He}} - \varepsilon_{\mathrm{H}}$$

where R is the approximate value—say that of hydrogen. If now M be the mass of a proton, that of the helium nucleus will be $3{\cdot}97M$. Hence

$$\varepsilon_{\mathrm{He}} = \frac{3{\cdot}97M}{3{\cdot}97M + m},$$

and

$$\varepsilon_{\mathrm{H}} = \frac{M}{M + m},$$

by **(40·681)**, therefore

$$\frac{R_{\mathrm{He}} - R_{\mathrm{H}}}{R} = \frac{m}{M} - \frac{m}{3{\cdot}97M}$$

therefore

$$\frac{R_{\mathrm{He}}' - R_{\mathrm{H}}'}{R'} = \frac{m}{M}\left(1 - \frac{1}{3{\cdot}97}\right)$$

in which R' signifies the value appropriate for giving wave-numbers. Now observation shows that

$$R_{\mathrm{He}}' - R_{\mathrm{H}}' = 44{\cdot}5 \text{ cm.}^{-1}.$$

Hence

$$\frac{44{\cdot}5}{109700} = \frac{m}{M}\left(1 - \frac{1}{3{\cdot}97}\right)$$

which gives us approximately

$$M/m = 1844,$$

in very satisfactory agreement with the value obtained by other and very different methods.

§ 40·8. Application of the Quantum Conditions to the Hydrogen Atom

According to the classical theory an electron in orbital motion round an attracting centre is continuously radiating energy on account of its acceleration **(§ 27·7).** This is a very slow process and to a first approximation such a system can be regarded as conservative. The application of the old quantum theory to such cases always assumed conservation except when transitions occurred from one stationary state to another. We shall therefore regard the system consisting of an electron in orbital motion round an attracting nucleus as subject only to a central force Ze^2/r^2, if r is the distance between the electron and the nucleus and ignore the reactionary forces due to the radiation of energy. We have already done this tacitly in describing Bohr's early work. The motion of the electron is therefore subject to the equation

$$\tfrac{1}{2}mv^2 - \frac{Ze^2}{r} = E, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(40·8)}$$

in which E is the energy, m and v are the mass and velocity respectively of the electron, e is the numerical value of the elementary charge ($e = 4{\cdot}78 \times 10^{-10}$ O.E.S. units approximately). We are assuming for brevity that the mass of the nucleus is very great, practically infinite. The reader will easily make for himself, with the aid of **§ 5·5** or **§ 40·6,** the necessary modifications for the finite nuclear mass. The motion of the electron is in a plane (cf. **§ 5·5**) and we may suppose this to be the X, Y plane. The energy, E, and the angular momentum, Ω, are constants of the motion (constants of integration). Other such constants are the **phase integrals**

$$\left.\begin{aligned} J_1 &= \oint p_r\, dr, \\ J_2 &= \oint p_\theta\, d\theta, \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad \textbf{(40·81)}$$

in which p_r and p_θ are the radial momentum and the angular momentum respectively, while r and θ are respectively the radius vector and the polar angle. These constants—the phase integrals **(40·81)**—can be expressed, as we shall see, in terms of E and Ω and the inherent constants of the system such as Z, m and e. We may in fact regard them as constants of integration which are alternative to E and Ω.

Equation **(40·8)** may of course be written

$$\frac{{p_r}^2}{2m} + \frac{{p_0}^2}{2mr^2} - \frac{Ze^2}{r} = E$$

and since p_θ is the constant represented by Ω,

$$p_r = \sqrt{\left|2mE + \frac{2mZe^2}{r} - \frac{\Omega^2}{r^2}\right|}.$$

When this is substituted in the first of the integrals **(40·81)** we get

$$J_1 = 2\pi\left\{\frac{mZe^2}{\sqrt{|-2mE|}} - \Omega\right\}.$$

For the method of effecting the integration the reader should consult the work of Sommerfeld mentioned in the bibliographical list at the end of this volume. We thus find

$$\left.\begin{aligned} E &= -\frac{2\pi^2 mZ^2e^4}{(J_1+J_2)^2}, \\ \Omega &= J_2/2\pi. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad \textbf{(40·82)}$$

The quantum conditions **(40·201)** require that

$$\left.\begin{aligned} J_1 &= n_1 h, \\ J_2 &= n_2 h, \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad \textbf{(40·83)}$$

where n_1 and n_2 are positive integers. Hence

$$E = -\frac{2\pi^2 mZ^2e^4}{n^2h^2}, \quad . \quad . \quad . \quad . \quad \textbf{(40·84)}$$

where n means $n_1 + n_2$. This is in agreement with **(40·62)** and, if we introduce the factor $\varepsilon(= M/(M+m))$, with **(40·68)**. It is now apparent that the validity of these formulae is not confined to circular orbits. The two equations **(40·82)** give us

$$-\frac{2E\Omega^2}{mZ^2e^4} = \frac{{J_2}^2}{(J_1+J_2)^2} = \frac{{n_2}^2}{(n_1+n_2)^2}$$

and it follows (by **5·532** and **5·551**) that

$$(1-e^2)^{1/2} = n_2/(n_1+n_2), \quad . \quad . \quad . \quad \textbf{(40·85)}$$

e being the eccentricity of the orbit.

§ 40·9. General Application of the Quantum Conditions

The application of the quantum conditions was confined, as we have seen, to systems which from the classical point of view are (*a*) conservative or so nearly so that they may be treated as such and (*b*) are such that co-ordinates q can be found with the property that each q_α librates between definite limits (as for example the r in the Kepler planetary motion) and each phase integral $\oint p_\alpha \, dq_\alpha$ is a constant. Such systems are called **conditionally periodic.** Each q_α passes periodically through the same range of values, but the periods of the different q's are not in general commensurable and the state of the *whole system* does not in general repeat itself. When it does happen that two or more periods are commensurable (as in the case of r and θ in Kepler motion) the system is said to be **degenerate** (cf. **§ 42·5**). Such constants of integration as the energy, angular momentum, etc., can be expressed in terms of the phase integrals J_α. Thus

$$E = function\ (J_1,\ J_2,\ J_3,\ \ldots) \quad . \quad . \quad . \quad \textbf{(40·9)}$$

and when we replace each J_α by $n_\alpha h$ we have an expression for the energy of the system in one of its stationary states. On dividing this by h (or ch) we get the associated spectral terms **(§ 40·6).** We may express the energy change ΔE, of the system, due to a transition from one stationary state to another, in the form

$$\Delta E = \Delta E_1 + \Delta E_2 + \Delta E_3 + \ldots$$

where ΔE_1 means the part of ΔE which is due to ΔJ_1 the change in J_1, and so on. Therefore

$$\Delta E = \frac{\Delta E_1}{\Delta J_1}\Delta J_1 + \frac{\Delta E_2}{\Delta J_2}\Delta J_2 + \ldots$$

If we divide both sides of this equation by h, the left-hand side is the frequency, ν, of the radiation emitted (or absorbed) during the transition and therefore

$$\nu = \frac{\Delta E_1}{\Delta J_1}s_1 + \frac{\Delta E_2}{\Delta J_2}s_2 + \ldots, \quad . \quad . \quad . \quad \textbf{(40·91)}$$

s_1, s_2, etc., being integers, positive, negative, or zero. This is the general expression, according to the older quantum theory, for the frequency of the radiation emitted by an atom.

We shall now show that the frequency of libration of the q_α,

assuming the validity of Newtonian (or Hamiltonian) dynamics, must be given by

$$\nu_\alpha = \frac{\partial E}{\partial J_\alpha}. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(40·92)}$$

The Hamiltonian function S is an invariant so that $\Sigma\int p_\alpha\, dq_\alpha$ is invariant and the dynamical systems in which we are now interested have the property (as explained above) that each separate $\int p_\alpha\, dq_\alpha$ is a function of q_α only and the integral $\oint p_\alpha\, dq_\alpha$ taken once over the complete range of libration of q_α is a constant—the constant represented by J_α. Now let us define a variable θ_α by

$$\theta_\alpha = \int p_\alpha\, dq_\alpha/\Omega_\alpha \quad . \quad . \quad . \quad . \quad . \quad \textbf{(40·93)}$$

in which, for the present, Ω_α may be any arbitrarily chosen constant. Thus S becomes

$$\Sigma\int \Omega_\alpha\, d\theta_\alpha - E\int dt.$$

θ_α is clearly a function of the single variable q_α and the Ω_α are new momenta. We shall fix the constant value of each Ω_α by laying down that

$$\oint p_\alpha\, dq_\alpha = \Omega_\alpha 2\pi = J_\alpha.$$

If the energy be expressed in the new canonical variables Ω_α and θ_α we have

$$\left.\begin{aligned} \frac{d\Omega_\alpha}{dt} &= -\frac{\partial E}{\partial \theta_\alpha}, \\ \frac{d\theta_\alpha}{dt} &= \frac{\partial E}{\partial \Omega_\alpha}. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad \textbf{(40·94)}$$

Since the Ω_α are constants the first of these equations indicates that the energy, E, can be expressed in terms of the Ω_α solely. Further, θ_α ranges through 2π while the variable q_α makes one whole vibration. The number of complete librations per unit time is $d\theta_\alpha/2\pi\, dt$. We shall represent this by ν, thus the frequency of libration of q_α is

$$\left.\begin{aligned} \nu_\alpha &= d\theta_\alpha/2\pi\, dt, \\ \text{or} \quad \nu_\alpha &= \partial E/2\pi\partial\Omega_\alpha. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad \textbf{(40·941)}$$

It is usual to represent $\theta_\alpha/2\pi$ by w_α and to call the latter

an angle variable. It increases by unity for each complete libration of q_α. Remembering that

$$J_\alpha = 2\pi\Omega_\alpha$$

we see that the equations **(40·94)** are equivalent to

$$\left.\begin{aligned} \frac{dJ_\alpha}{dt} &= -\frac{\partial E}{\partial w_\alpha} \\ \frac{dw_\alpha}{dt} &= \frac{\partial E}{\partial J_\alpha} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad \textbf{(40·95)}$$

the latter of these two equations being equivalent to **(40·92)**. As a verification let us take the case, already discussed, of the circular motion of an electron round a massive positively charged nucleus. The energy E is expressed by

$$-E = \frac{2\pi^2 m Z^2 e^4}{J^2}$$

where

$$J = mv2\pi r.$$

The frequency of revolution, according to **(40·92)**, is

$$\nu = \frac{4\pi^2 m Z^2 e^4}{J^3} = \frac{-2E}{J}$$

or

$$\nu = \frac{2T}{J}$$

i.e.

$$\nu = \frac{mv^2}{mv2\pi r} = \frac{v}{2\pi r}$$

and the formula is verified.

A further and helpful illustration is provided by simple harmonic motion—a particle of mass m moving under the influence of a restoring force equal to μ (a constant) per unit displacement. Obviously

$$\frac{p^2}{2m} + \frac{\mu}{2}q^2 = E$$

and consequently

$$p = \sqrt{2mE - m\mu q^2}$$

or

$$p = \sqrt{m\mu}\sqrt{\frac{2E}{\mu} - q^2}$$

Hence since

$$p\,dq = \Omega\,d\theta$$

$$2\pi\Omega = \oint \sqrt{m\mu}\sqrt{\frac{2E}{\mu} - q^2}\,dq.$$

Therefore

$$2\pi\Omega = \sqrt{m\mu\,|} \times \left(\text{Area of circle of radius } \sqrt{\frac{2E}{\mu}\Big|}\right)$$

$$2\pi\Omega = \sqrt{m\mu\,|} \times \pi\frac{2E}{\mu}$$

i.e. $$J = 2\pi\sqrt{\frac{m}{\mu}\Big|}.E.$$

Here again the classical frequency, which is $\sqrt{\mu/m\,|}\,/2\pi$, is equal to $\partial E/\partial J$.

The frequencies **(40·92)** must appear (according to the classical theory) in the radiation emitted by conditionally periodic systems. It is easy to see, however, that they are in general only the fundamental frequencies and that integral multiples of them must generally appear. Let us consider, for example, the disturbance produced in the surrounding medium in consequence of the periodic change of r, the radial distance of the electron from the centre of mass in the hydrogen-like atom. Though periodic, the variation is not a simple harmonic one and the spectroscope carries out a Fourier analysis of it into its simple harmonic components. Thus we come to the conclusion that the frequencies of the emitted radiation *according to the classical theory* must be expressed by

$$\nu = \frac{\partial E}{\partial J_1}s_1 + \frac{\partial E}{\partial J_2}s_2 + \frac{\partial E}{\partial J_3}s_3 + \ldots, \qquad \textbf{(40·96)}$$

the s_1, s_2, s_3, etc., being integers positive, negative, or zero.

There is a remarkable resemblance between this classical formula and the quantum formula **(40·91)**. To each frequency as it might be calculated by classical methods there is a corresponding one given by **(40·91)**. Apart from the fact that we have in **(40·91)** ratios of finite differences and differential quotients in **(40·96)**, there is another important distinction between the two formulae. In the classical formula **(40·96)** some of the integers, s, have quite definite values assigned to them while others may have any integral value. We can illustrate this by a simple harmonic motion which is characterized by one frequency, namely,

$$\nu = \frac{\partial E}{\partial J} = \frac{1}{2\pi}\sqrt{\frac{\mu}{m}},$$

so that s is unity. Another illustration can be found in the case of the s associated with $\partial E/\partial J$ where J is the angular phase integral in the Kepler motion. Here s is confined to the values $+$ or -1 (cf. **§ 42·8**).

§ 41. The Correspondence Principle

We can now understand the basis and significance of Bohr's famous **correspondence principle.** The classical formula **(40·96)** is clearly an asymptotic limit to which the quantum formula **(40·91)** approaches as the scale of magnitude of the quantities involved becomes very great, in particular when the ΔE_α and ΔJ_α become very small compared with E_α and J_α respectively. Now we have seen **(§ 40·5)** that, while every spectral frequency is equal to the difference between two of a set of terms characteristic of the emitting (or absorbing) system the converse is not always the case. We have seen, for example, that in series spectra the frequencies (or wave-numbers) may be differences of S and P terms, or of P and D terms but *not* of S and D terms. Bohr made some progress in introducing rationality into these restrictions (**selection rules** they were called) and also in other directions, e.g. the determination of the polarization of the components in the Stark resolution of a spectral line (cf. **§ 41·1**) by applying his **correspondence principle.** This simply lays down that those things which the quantum theory (without the correspondence principle) left undetermined should be determined by the aid of the corresponding classical formula or expression. If, for example, the classical theory gives a definite value, or set of values, for one of the integers s in **(40·96)** the correspondence principle prescribes this value (or values) for the corresponding s in **(40·91).** Or when, for example, the classical theory indicates a particular type of polarization, the correspondence principle lays down the same type of polarization for the quantum theoretical radiation of the corresponding frequency. Similarly, when the classical theory associates a definite intensity with the emitted radiation of some frequency, Bohr's correspondence principle assigns the same relative intensity to the radiation which has the corresponding frequency as given by **(40·91).** In simple harmonic systems $s = 1$ for example. Similarly in the Kepler motion the integer s_2 is confined to the values $+1$ or -1. In other words, the integer n_2 (Bohr's k) must change by $+1$ or -1 when a transition occurs. One of Bohr's strokes of genius consisted in assigning to k the values 1, 2, 3, etc., for the S, P, D, F . . . terms, respectively (cf. **§ 40·5**), thus giving some reason for the absence of such frequencies as that represented by $S - D$, for example, since $k(= n_2)$ can only change by $+1$ or -1.

Further illustrations of the application of the correspondence principle will appear in the sections devoted to the Stark and Zeeman phenomena. It is really only an aspect of a wider

correspondence, a limiting case of which has long been known and consists in the perfect analogy between geometrical optics and Hamiltonian dynamics (cf. **§§ 9·4, 41·5** and **41·6**).

§ 41·1. The Stark Effect

When emitting hydrogen atoms are subjected to a strong external electric field each spectral line (which for the present purpose may be considered to be single since the effect we are now going to study is on a much bigger scale than that of the fine structure which the line exhibits under highly resolving apparatus) splits up into a group of lines which are symmetrically situated relatively to the position of the original line. The phenomenon is named after Johannes Stark, who discovered it in 1913. Its explanation by the application of the quantum conditions (cf. **§§ 40·8** and **40·9**) amplified by the correspondence principle was one of the great successes of the older quantum theory. The mathematical problem involved is a particular case of the problem of the motion of a particle under the attractive forces of two fixed centres of force and was solved by the German mathematician C. G. J. Jacobi in the earlier half of last century.

We shall suppose the attracting nucleus to have the charge Ze and to be at the origin of rectangular co-ordinates. Superposed on the field of the nucleus is a uniform one of intensity F which we shall assume to be in the X direction. In order that the variables may be separated in the sense explained in **§ 40·9** we choose co-ordinates ξ and η (known as **parabolic co-ordinates**) and defined by

$$x = (\xi^2 - \eta^2)/2, \quad . \quad . \quad . \quad . \quad . \quad (41\cdot1)$$
$$\rho = \xi\eta,$$

where x is the X co-ordinate of the planetary electron and $\rho[= (y^2 + z^2)^{1/2}]$ is its perpendicular distance from the X axis. For the remaining co-ordinate we choose the azimuthal angle, ϕ, about the X axis. We shall content ourselves with giving the expression for the energy in terms of the phase integrals J_ξ, J_η, and J_ϕ, namely

$$E = -\frac{2\pi^2 m Z^2 e^4}{J^2} - \frac{3}{8}\frac{J(J_\xi - J_\eta)}{\pi^2 m Ze}, \quad . \quad . \quad (41\cdot11)$$

where $$J \equiv J_\xi + J_\eta + J_\phi.$$

In arriving at the formula it has been assumed that the centre of mass of the atom is in the nucleus itself an assumption which is of course approximately true.

Equation **(41·11)** does not involve any departure from classical dynamical methods.

The quantum theory enters when we replace each J by the product of an integer and h, for example

$$J_\xi = n_\xi h,$$

n_ξ being an integer, positive or zero, and when we identify E/h with a spectral term (or E/ch if wave-numbers are being calculated). Thus we get for a spectral term

$$-\frac{E}{h} = +\frac{2\pi^2 mZ^2e^4}{n^2h^3} + \frac{3}{8}\frac{n(n_\xi - n_\eta)h}{\pi^2 mZe}F, \quad . \quad . \quad \textbf{(41·12)}$$

where $$n \equiv n_\xi + n_\eta + n_\phi.$$

In place, therefore, of the single term $2\pi^2 mZ^2e^4/n^2h^3$ associated with a given value of n, we have a number of terms.

The correspondence principle restricts the transitions to those for which $\Delta n_\phi = +1$, 0, or -1. Moreover, it indicates that transitions for which $\Delta n_\phi = 0$ are associated with radiation polarized in the sense that the electric vibrations in it are parallel to the external field F (parallel to X) and those transitions for which $\Delta n_\phi = +1$ or -1 are associated with radiation which is circularly polarized as viewed along the axes of F. Both radiations will appear plane polarized as viewed perpendicularly to F, the former having its electric vibrations parallel to F and the latter perpendicular. In the case of the circularly polarized radiations both directions of rotation for any given frequency are equally frequent, so that the radiation as actually observed longitudinally, coming as it does from many atoms, appears unpolarized.

Not only are the transitions confined to those for which $\Delta n_\phi = +1$, 0, or -1, but n_ϕ cannot be zero. The correspondence principle does appear to indicate this. The integer n_ϕ corresponds to the integer $l + 1$ (cf. **§§ 42·3, 42·4,** and **42·5**) which cannot be zero.

The differences in frequency between the components, into which a line (say a Balmer line) is broken up, are represented by the expression

$$\left.\begin{aligned} \Delta\nu &= \frac{3}{8}\frac{hF}{\pi^2 mZe}n(n_\xi - n_\eta), \\ \text{or} \qquad \Delta\nu &= CN \end{aligned}\right\} \quad . \quad . \quad . \quad \textbf{(41·13)}$$

where C means $3hF/8\pi^2 mZe$ and N means $n(n_\xi - n_\eta)$. The separations are therefore proportional to the field F.

We shall illustrate **(41·13)** by applying it to the special case of the Balmer α line of hydrogen. The atomic number, Z, is equal to unity. Therefore $C = 3hF/8\pi^2 me$. The α line is due

to a transition from $n = 3$ to $n = 2$. The possibilities for the separate quantum integers are given in the following table:

Before Transition				After Transition			
n_ϕ	n_ξ	n_η	N	n_ϕ	n_ξ	n_η	N
3	0	0	0	2	0	0	0
2	1	0	3	1	1	0	2
2	0	1	− 3	1	0	1	− 2
1	2	0	6				
1	0	2	− 6				
1	1	1	0				

Thus we get for the p components, i.e. those polarized with their vibrations parallel to the field ($\Delta n_\phi = 0$):

When $n_\phi = 2$

$$(3 - 0)C = 3C$$
$$(-3 - 0)C = -3C$$

When $n_\phi = 1$

$$(6 - 2)C = 4C$$
$$(-6 - 2)C = -8C$$
$$(0 - 2)C = -2C$$
$$(6 + 2)C = 8C$$
$$(-6 + 2)C = -4C$$
$$(0 + 2)C = 2C$$

For the s components, polarized perpendicularly to the direction of the field ($\Delta n_\phi = +1$ or -1):

Initial $n_\phi = 3$

$$(0 - 0)C = 0C$$

Initial $n_\phi = 2$

$$(3 - 2)C = C$$
$$(3 + 2)C = 5C$$
$$(-3 - 2)C = -5C$$
$$(-3 + 2)C = -C$$

Initial $n_\phi = 1$

$$(6 - 0)C = 6C$$
$$(-6 - 0)C = -6C$$
$$(0 - 0)C = 0C$$

The components are therefore symmetrically situated, their separations being integral multiples of the quantity C. The multiples for the p components being 2, 3, 4, and 8 on one side

and -2, -3, -4, and -8 on the other. Those of the s components are

0, 1, 5, and 6

with -1, -5, and -6.

The agreement with Stark's observations is extraordinarily good, even in regard to the relative intensities of the components, which have been estimated by Kramers with the aid of the correspondence principle. One very noteworthy feature is that the theory makes the p and s components of H_β to coincide and thus predicts partial polarization which is in fact what is observed.

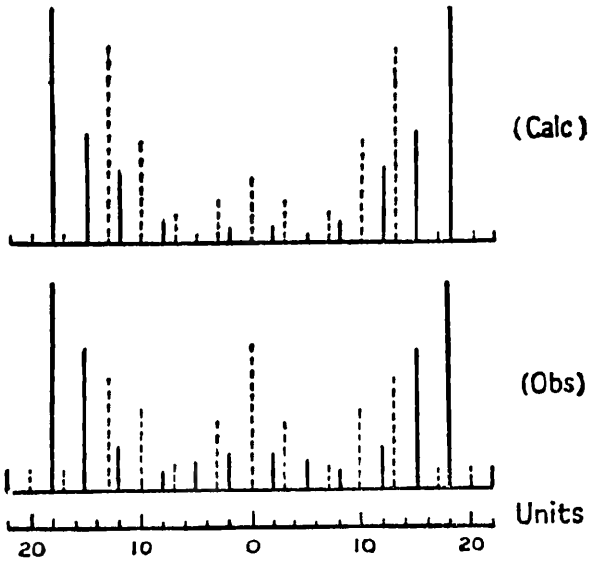

FIG. 41·1

In the figure [1] the p lines are drawn in full, the s lines are broken. The heights of the vertical lines in the figure represent intensities. The agreement between the observed results (lower part of figure) and the theoretically calculated ones (upper part) is very striking.

§ 41·2. THE ZEEMAN EFFECT

The simple Zeeman effect can be accounted for, as H. A. Lorentz showed (**§ 26·7**), without resort to the quantum theory by supposing that the radiation of the particular spectral line is due to the simple harmonic vibrations of an ion (electron).

Precisely the same result emerges when we apply the classical theory to the hydrogen-like atom. The problem is so closely linked with the recent development of quantum mechanics that it is desirable to deal with it here. We have learned in **§ 36·1** that the equations of motion of a charged particle in an electromagnetic field are of the same form as those for a particle in the absence of such a field provided we replace the momentum $\mathbf{p}$ by a more general momentum $\mathbf{\Pi}$ defined by

$$\mathbf{\Pi} = \mathbf{p} + e\mathbf{K}. \quad . \quad . \quad . \quad . \quad . \quad (41\cdot2)$$

We may call $\mathbf{K}$ the electromagnetic potential. Its **curl** is equal to $\mu\mathbf{H}/a$, where a is the numerical constant whose value is unity when we employ electromagnetic or electrostatic units and equal

[1] The figure is taken for R. C. Johnson's Spectra and illustrates the case of H_γ.

to c when we use mixed units—electromagnetic for the magnetic quantities and electrostatic for the electrical quantities. We shall use the latter units and may take μ to be unity. Thus

$$\mathbf{curl}\ \mathbf{K} = \mathbf{H}/c. \quad . \quad . \quad . \quad . \quad . \quad (41{\cdot}21)$$

It should be noted that in **(41·2)** the symbol e includes the sign of the charge as well as its absolute value. Thus for an electron it means $-4{\cdot}78 \times 10^{-10}$ O.E.S. It is, however, convenient in what follows to use e to mean $+4{\cdot}78 \times 10^{-10}$ in these units. Let us now consider the motion of an electron in the field of an attracting nucleus whose charge is $+Ze$. The application to it of a magnetic field, **H**, will not change its energy (since the resulting contribution to the force on the electron will be perpendicular to the direction of its motion); but the motion will be modified in the sense that instead of the azimuthal angular momentum, p_ϕ about the axis of **H**, remaining constant, the momentum

$$\Pi_\phi = p_\phi - eK_\phi$$

will be constant, where K_ϕ is a vector generalized in the same sense as p_ϕ.

Now **(41·21)** is satisfied by

$$K_x = -yH/2c,$$
$$K_y = xH/2c,$$

and

$$K_z = 0,$$

H being identical with H_z. These equations, it will be observed, satisfy

$$\mathbf{div}\ \mathbf{K} = 0.$$

The vector **K** is parallel to planes perpendicular to **H** and

$$\mathbf{K}^2 = r^2 \sin^2\theta . \mathbf{H}^2/4c^2$$

or

$$\mathbf{K} = r \sin\theta . \mathbf{H}/2c.$$

The appropriate expression for K_ϕ is clearly

$$K_\phi = r \sin\theta . \mathbf{K} = r^2 \sin^2\theta . \mathbf{H}/2c.$$

The field **H** is in the direction of the Z axis and θ is the angle between this direction and the radius vector, r.

We now have

$$\Pi_\phi \equiv p_\phi - eK_\phi = p_\phi - r^2 \sin^2\theta . e\mathbf{H}/2c = \mathbf{M}, \text{ a constant.} \quad (41{\cdot}22)$$

The energy equation is

$$\frac{p_r^2}{2m} + \frac{p_\theta^2}{2mr^2} + \frac{p_\phi^2}{2mr^2 \sin^2\theta} - \frac{Ze^2}{r} = E, \quad . \quad . \quad (41{\cdot}23)$$

or, by **(41·22)**

$$\frac{p_r^2}{2m} + \frac{p_\theta^2}{2mr^2} + \frac{(\mathbf{M} + r^2 \sin^2\theta . e\mathbf{H}/2c)^2}{2mr^2 \sin^2\theta} - \frac{Ze^2}{r} = E.$$

Therefore, if we ignore the second order small term containing $\mathbf{H}^2$,

$$p_\theta^2 + \frac{\mathbf{M}^2}{\sin^2\theta} = \left(2mE - \frac{\mathbf{MeH}}{c} - p_r^2\right)r^2 + 2mZe^2r. \quad . \quad (41{\cdot}231)$$

Each side of this equation must be equated to a constant, which we shall represent by N. Thus the phase integrals are:

$$J_r = \oint \Pi_r\, dr = \oint p_r\, dr$$

$$= \oint \sqrt{\left\{\frac{-N}{r^2} + \frac{2mZe^2}{r} + \left(2mE - \frac{\mathbf{MeH}}{c}\right)\right\}} dr$$

$$J_\theta = \oint \Pi_\theta\, d\theta = \oint p_\theta\, d\theta$$

$$= \oint \sqrt{\left\{N - \frac{\mathbf{M}^2}{\sin^2\theta}\right\}} d\theta,$$

and

$$J_\phi = \oint \Pi_\phi d\phi = 2\pi\mathbf{M}.$$

Thus

$$\left.\begin{aligned} J_r &= 2\pi\left\{\frac{mZe^2}{\sqrt{\left|\frac{\mathbf{MeH}}{c} - 2mE\right|}} - \sqrt{N}\right\}, \\ J_\theta &= 2\pi\{\sqrt{N} - \mathbf{M}\}, \\ J_\phi &= 2\pi\mathbf{M}. \end{aligned}\right\} \quad . \quad . \quad (41{\cdot}24)$$

On eliminating the constants $\mathbf{M}$ and N from these equations we easily get

$$J_r + J_\theta + J_\phi = \frac{2\pi mZe^2}{\sqrt{\left\{\frac{J_\phi e\mathbf{H}}{2\pi c} - 2mE\right\}}}.$$

Hence

$$\frac{J_\phi e\mathbf{H}}{2\pi c} - 2mE = \frac{4\pi^2 m^2 Z^2 e^4}{(J_r + J_\theta + J_\phi)^2},$$

or

$$-E = \frac{2\pi^2 mZ^2e^4}{(J_r + J_\theta + J_\phi)^2} - J_\phi \frac{e\mathbf{H}}{4\pi mc}. \quad . \quad . \quad . \quad (41{\cdot}25)$$

We get the possible frequencies, according to the classical theory, by applying the formula **(40·92)**. The field $\mathbf{H}$ does not affect the frequencies associated with J_r and J_θ; but that associated with J_ϕ is changed by the amount

$$\Delta\nu = \frac{e\mathbf{H}}{4\pi mc}.$$

Now p_ϕ may be positive or negative as referred to the Z axis (direction of **H**), i.e. it may be in the sense of the rotation of a screw travelling in the Z direction (p_ϕ positive) or the opposite sense. In the former case $\Delta\nu$ is an addition to ν and in the latter a reduction of ν. So we have the original frequency ν and two new frequencies $\nu + \Delta\nu$ and $\nu - \Delta\nu$, where $\Delta\nu = \dfrac{e\mathbf{H}}{4\pi mc}$.

We might have arrived at this result much more simply. Indeed, the application of the field **H**, according to **(41·22)**, increases p_ϕ from its original value **M** to $\mathbf{M} + eK_\phi$. The associated increase of angular velocity is

$$eK_\phi/mr^2 \sin^2\theta = e\mathbf{H}/2mc.$$

Thus the effect of the field is to superpose on the angular velocity $d\phi/dt$ the amount $e\mathbf{H}/2mc$. This effect is called the **Larmor precession.** In other respects the motion is changed inappreciably, since in the energy equation, as represented by **(41·231)**, the constant $2mE$ is altered only by the small term $\mathbf{M}e\mathbf{H}/c$. In fact, we may describe the effect of the field **H** by saying that it causes the ellipse, in which the electron moves when $\mathbf{H} = 0$, to precess round the Z axis (axis of **H**) with the Larmor precessional velocity given above. It will be seen in fact that if we were to refer the system to co-ordinates rotating about Z (or the axis of **H**) with the angular velocity $e\mathbf{H}/2mc$ the electron will travel in its usual elliptical orbit (or, to be quite precise, in the orbit which is modified by replacing E by $E - \mathbf{M}e\mathbf{H}/2mc$ which is inappreciably different from E). The angle between the normal to the plane of the precessing ellipse and the direction of **H** is thus constant, so that the projection of the radius vector, r, on the axis of **H** is unmodified by the field and obviously it is not affected by the Larmor precession. On the other hand, the projection of r on a plane perpendicular to **H** is affected by the Larmor precession. We conclude, therefore, that the radiation which would be emitted in directions perpendicular to **H** if the classical theory were valid will be polarized with its vibrations parallel to **H** in the case of the unmodified frequency, ν, and with its vibrations perpendicular to **H** in the case of the frequencies $\nu \pm \Delta\nu$. The Larmor precession, it is clear, will impress on the radiation travelling in the line of **H** a circular polarization—not a pure circular polarization since the projection of the electron's motion on a plane perpendicular to **H** is not in general a circle—in such a sense that the higher frequency is associated with a clockwise motion of the electron when it is seen by looking in the direction of **H.**

To apply the quantum theory to the problem we have to

replace each of the phase integrals J by the product of an integer and h and form the spectral terms by dividing $-E$ by h. We thus get for the spectral terms:

$$-\frac{E}{h} = \frac{2\pi^2 m Z^2 e^4}{n^2 h^3} - n_\phi \frac{eH}{4\pi mc}$$

where $n = n_r + n_\theta + n_\phi$. Thus the old quantum theory also yields

$$\Delta\nu = \frac{eH}{4\pi mc} s_\phi$$

where, in accordance with the correspondence principle,

$$s_\phi = \Delta n_\phi = +1,\ 0,\ \text{or}\ -1,$$

the changes $+1$ and -1 being associated with circularly polarized light. Thus the old quantum theory gives the same pronouncements about the Zeeman effect as does the classical theory. The integer n_ϕ has been called by Sommerfeld the **magnetic quantum number.**

CHAPTER X

INTRODUCTION TO WAVE-MECHANICS

§ 41·3. A Certain Fourier Expansion

THE function $f(z)$, defined by

$$f(z) = 0\begin{cases} -L < z < -l, \\ l < z < L, \end{cases}$$

and

$$f(z) = A \cos \Omega z, \quad -l < z < l,$$

where z is a real variable and l, L, A, and Ω are real and positive constants, can be expanded in the form

$$f(z) = a_0 + a_1 \cos \frac{\pi}{L} z + a_2 \cos \frac{2\pi}{L} z + a_3 \cos \frac{3\pi}{L} z + \ldots \quad (41{\cdot}3)$$

The coefficients, a, are given by the formulae

$$\left.\begin{aligned} a_0 &= \frac{1}{2L}\int_{-L}^{L} f(\zeta)\, d\zeta, \\ a_s &= \frac{1}{L}\int_{-L}^{L} f(\zeta) \cos \frac{s\pi}{L}\zeta \; d\zeta, \end{aligned}\right\} \quad \ldots \quad (41{\cdot}31)$$

where $s = 1, 2, 3, \ldots \infty$. Sine terms are absent in the expansion, since $f(z)$ is an even function **(§ 4)**. On carrying out the integrations in **(41·31)** we easily find

$$\left.\begin{aligned} a_0 &= \frac{A}{L}\frac{\sin \Omega l}{\Omega}, \\ a_s &= \frac{A}{L}\left\{\frac{\sin\left(\Omega + \frac{s\pi}{L}\right)l}{\Omega + \frac{s\pi}{L}} + \frac{\sin\left(\Omega - \frac{s\pi}{L}\right)l,}{\Omega - \frac{s\pi}{L}}\right\} \end{aligned}\right\} \quad (41{\cdot}311)$$

where again $s = 1, 2, 3, \ldots \infty$. For convenience, let us call those quantities which are of the same order of magnitude as unity quantities of the order **null**. Those which are very large compared with unity we shall call quantities of the order **+ 1**.

Quantities of the order magnitude of the squares or products two at a time of those which are of the order + 1 we shall term quantities of the order + 2, and so on. Thus if we were to adopt 10^6 as a number of the order + 1, then 10^{12} would be one of the order + 2. Similarly we shall speak of quantities which are small compared with unity as quantities of the order − 1; those of the order of magnitude of squares or products two at a time of quantities of the order − 1 we shall call quantities of the order − 2, and so on. Using the standard we have adopted as an illustration, 10^{-6} and 10^{-12} might be regarded as of the orders − 1 and − 2 respectively.

We are interested in the case where

$$\left.\begin{array}{lll} A & \text{is of the order} & 0, \\ l & ,, \quad ,, & 0, \\ \Omega & ,, \quad ,, & +1, \\ L & ,, \quad ,, & +1, \end{array}\right\} \quad . \quad . \quad . \quad \textbf{(41·32)}$$

and we shall regard the standard which fixes the order + 1 as a very large number, so large that it may be taken for practical purposes as infinite.

The **wave-length,** λ, i.e. the distance from crest to crest, in the cosine curve described by $f(z)$ is obviously equal to $2\pi/\Omega$ and we shall define a number, σ (not necessarily an integer), by

$$\Omega = \sigma\pi/L. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(41·321)}$$

It will also be convenient to use Ω_s in the sense defined by

$$\Omega_s = s\pi/L, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(41·322)}$$

where the s are the integers in the expansion **(41·3)** and in **(41·31).** If we use x and y to mean respectively $\Omega_s - \Omega$ and $\Omega_s + \Omega$,

$$x = (s - \sigma)\pi/L$$
$$y = (s + \sigma)\pi/L$$

and we may write

$$dx = dy = \pi/L.$$

The sum **(41·3)** may therefore be written

$$f(z) = \frac{A}{\pi}\left\{\int_{\Omega}^{\Omega+\infty} \cos\,(y - \Omega)z\,\frac{\sin yl}{y}dy + \int_{-\Omega}^{\infty} \cos\,(x + \Omega)z\,\frac{\sin xl}{x}dx,\right\}$$

The integral in y, both limits of which are positive and enormously large, must vanish and consequently

$$f(z) = \frac{A}{\pi}\int_{-\Omega}^{+\infty} \cos\,(x + \Omega)z\,\frac{\sin xl}{x}dx. \quad . \quad . \quad \textbf{(41·323)}$$

Equations **(41·3)** and **(41·323)** represent $f(z)$ as a superposition of cosine curves of wave-lengths extending from infinity ($\Omega_s = 0$) to zero ($\Omega_s = \infty$), corresponding in the latter equation to a range of values of x from $-\infty$ to $+\infty$. Now it is obvious from **(41·311)** that the only amplitudes which are important in this superposition are those corresponding to small values of $(\Omega_s - \Omega)/\Omega$, i.e. to a small range of values of x in the integral **(41·323)** and it is of interest to inquire about the result of a superposition of cosine curves for all of which $(\Omega_s - \Omega)/\Omega$ is very small, i.e. curves for which $\Delta\left(\frac{1}{\lambda_s}\right)$ is small compared with $1/\lambda$. We therefore ask ourselves: What is the value of the integral

$$\frac{A}{\pi}\int_{-a}^{+a}\cos(x+\Omega)z\,\frac{\sin xl}{x}dx, \quad . \quad . \quad . \quad \textbf{(41·33)}$$

where a is a small positive number? So long as the extreme values of xz are sufficiently small we may write for this integral

$$\frac{A}{\pi}\cos\Omega z\int_{-a}^{+a}\frac{\sin xl}{x}dx$$

which becomes, when l is sufficiently large, $A\cos\Omega z$. This will be approximately true even if l is not infinite. When l is sufficiently large, therefore, there is an appreciably large range of values of z, say $-\Lambda < z < +\Lambda$, for which $a\Lambda$ is so small that $\cos(x+\Omega)z$ does not differ sensibly from $\cos\Omega z$. This is no longer the case when $a\Lambda$ becomes an appreciable fraction of $\pi/2$, so that **(41·33)** will only represent the function $A\cos\Omega z$ within the range of values $-\Lambda < z < \Lambda$, when the order of magnitude of $a\Lambda$ does not exceed that of unity, so that

$$a\Lambda \sim 1$$

at most, or

$$\Delta\Omega \times \Lambda \sim 1,$$

and since $\Delta\Omega = 2\pi\Delta\left(\frac{1}{\lambda}\right)$,

$$2\pi\Delta\left(\frac{1}{\lambda}\right) \times \Lambda \sim 1$$

or finally

$$\boxed{\Delta(1/\lambda) \times 2\Lambda \sim 1} \quad . \quad . \quad . \quad . \quad \textbf{(41·34)}$$

When z is outside this range of values, i.e. when $z > \Lambda$ or $z < -\Lambda$, the function $\cos(x+\Omega)z$ in the integral **(41·33)** begins to differ appreciably from $\cos\Omega z$ and, when z is sufficiently

far outside the range $-\Lambda < z < +\Lambda$, to alternate in sign so that the integral **(41·33)** becomes small and rapidly approaches zero the farther z extends beyond the limits marked by Λ and $-\Lambda$

We arrive at the conclusion therefore that a superposition of cosine curves extending from $-\infty$ to $+\infty$, and all of nearly the same wave-length, will produce a cosine curve of wave-length λ, differing little from the wave-lengths of any of the constituent cosine curves, within certain limits $-\Lambda < z < +\Lambda$. Beyond these limits it will gradually and fairly rapidly approach zero as illustrated in Fig. 41·3.

FIG. 41·3

§ 41·4. THE SIMPLE GROUP

The function $f(z)$ of **§ 41·3,** which was defined between the limits

$$-l < z < l$$

to be

$$f(z) = A \cos \Omega z,$$

will become, when we replace z by $z' - d$,

$$\psi(z') = A \cos \Omega(z' - d) \quad . \quad . \quad . \quad . \quad \textbf{(41·4)}$$

between the limits

$$-l + d < z' < l + d.$$

Outside these limits it vanishes. The centre of the wave system which $f(z)$ or $\psi(z')$ describes is at $z = 0$, or at $z' = d$. If we replace d by ut, where u is a constant velocity of the order 0 and t is the time, and if we then write z for z'—this will cause no confusion since we shall have no further need for z in its original sense—formula **(41·4)** becomes

$$\psi = A \cos \Omega(z - ut) \quad . \quad . \quad . \quad . \quad \textbf{(41·41)}$$

and we conclude that the train of waves it describes can be regarded as a superposition of simple waves

$$\psi = \Sigma a_s \cos \Omega_s(z - ut), \quad . \quad . \quad . \quad \textbf{(41·42)}$$

each simple sinusoidal wave in the superposition **(41·42)** extending from $-\infty$ to $+\infty$, and if we ignore in the summation the terms for which s lies outside the range Δs defined in **§ 41·3** it will still approximate to $A \cos \Omega(z - ut)$ when

$$-\Lambda + ut < z < \Lambda + ut. \quad . \quad . \quad . \quad \textbf{(41·421)}$$

We shall use the term **simple group** for this train of waves and it will be noted that, in this instance, the group travels forward with the same velocity, namely, u, as that of the individual crests and troughs. In other words, the **group velocity** is identical, in the present example, with the **phase velocity.**

Let us next study a simple group or superposition of simple waves like that represented by **(41·42),** but with the difference that each constituent simple wave has its own individual phase velocity. We must describe it by

$$\psi = \sum_s a_s \cos \Omega_s(z - u_s t), \quad . \quad . \quad . \quad \textbf{(41·43)}$$

or

$$\psi = \sum_s a_s \cos (\Omega_s z - T_s t), \quad . \quad . \quad . \quad \textbf{(41·431)}$$

where T_s means the product $\Omega_s u_s$. Each u_s in the narrow range Δs defined in **§ 41·3** differs very little, we shall suppose, from a certain mean velocity, u, and we shall also assume that the ratio

$$(u_s - u)/(\lambda_s - \lambda) \text{ or } du/d\lambda$$

is sensibly constant within the narrow range Δs or $\Delta\lambda$ with which we are concerned. Now it is clear that at the instant $t = 0$ the superposition of simple waves described by **(41·43)** is identical with that described by **(41·42).** It will also be identical with it at all the instants

$$t = nP$$

where n is an integer, positive or negative, and

$$P = d\lambda/du,$$

or

$$P = -d\lambda/du,$$

according as $d\lambda/du$ is positive or negative. We can see this with the help of the conventionalized diagram, Fig. 41·4. At the instant $t = 0$ crests of all the individual simple waves will be coincident at O. The crests which are one wave-length to the

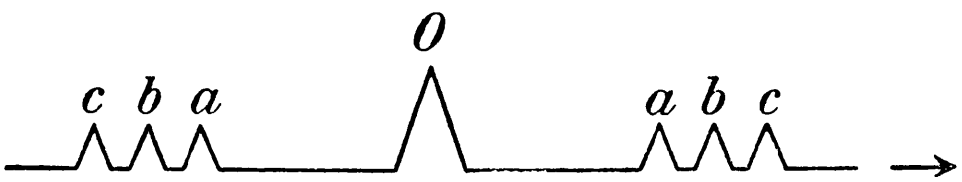

FIG. 41·4

left, or to the right, of O will occupy separate positions at a, b, c, . . . If now, for example, the simple waves of longer wave-length travel faster than those of shorter wave-length, the crest c on the left will overtake the crest a in the time

$$(\lambda_c - \lambda_a)/(u_c - u_a),$$

i.e. in the time $P = d\lambda/du$. Similarly the crest b will overtake

the crest a in the time $(\lambda_b - \lambda_a)/(u_b - u_a)$, i.e. in the time $P = d\lambda/du$. After the time, P, therefore the centre of the group has dropped a distance λ behind O, or the centre of the group has travelled the distance $uP - \lambda$. Its velocity is therefore

$$v = (uP - \lambda)/P$$

or

$$v = u - \lambda/P$$

or, finally

$$v = u - \lambda\, du/d\lambda. \quad . \quad . \quad . \quad . \quad \textbf{(41·44)}$$

When $d\lambda/du$ is negative it is easy to see that the crest a, on the right for example, will overtake the crests b and c in the time $P = -d\lambda/du$, and during this time the centre of the group has consequently travelled the distance $uP + \lambda$ and just as before the **group velocity** will be that given by **(41·44)**.

We see that at the instants $t = nP$ the superposition of simple waves described by **(41·43)** has exactly the same shape as that described by **(41·42)**. What can we say about its shape at intermediate times? To answer this question we replace z by $\zeta + vt$, in which v is the group velocity, i.e. the velocity of the centre of the group, so that ζ means distances measured from the centre of the group. Thus

$$\psi = \sum_s a_s \cos\{\Omega_s(\zeta + vt) - T_s t\}$$

or

$$\psi = \sum_s a_s \cos\{\Omega_s\zeta - (T_s - v\Omega_s)t\}.$$

The phase velocity of the simple wave, s, *relatively to the centre of the group*, is thus

$$T_s/\Omega_s - v \text{ or } u_s - v,$$

where u_s is its phase velocity relative to the fixed zero $z = 0$. We have now a clear picture of the group. It has the shape of a sine curve, approximately at any rate; the group as a whole progresses with the velocity v given by **(41·44)**. The individual crests or troughs travel through it with a velocity $u - v$ relative to the group, being created behind and destroyed in front when $u - v$ is positive ($d\lambda/du$ positive) or destroyed behind and created in front when $u - v$ is negative.

There is another way of expressing the formula **(41·43)** which is interesting and important. We may write it

$$\psi = \sum_s a_s \cos\{(\Omega + \Delta\Omega_s)z - (T + \Delta T_s)t\},$$

or

$$\psi = \sum_s a_s \cos\{(\Omega z - Tt) + (\Delta\Omega_s z - \Delta T_s t)\},$$

and therefore

$$\psi = \cos(\Omega z - Tt)\sum_s a_s \cos(\Delta\Omega_s z - \Delta T_s t) - \sin(\Omega z - Tt)\sum_s a_s \sin(\Delta\Omega_s z - \Delta T_s t)$$

and within the range Δs already defined this must be

$$\psi = A \cos (\Omega z - Tt),$$

so that

$$\left.\begin{aligned} A &= \Sigma a_s \cos (\Delta\Omega_s z - \Delta T_s t), \\ 0 &= \Sigma a_s \sin (\Delta\Omega_s . z - \Delta T_s . t). \end{aligned}\right\} \quad . \quad . \quad . \quad \textbf{(41·45)}$$

Strictly speaking, A represented by **(41·45)** is not *quite* constant. It varies from a maximum at the centre of the group, changes very slowly within the limits **(41·421)** and rapidly approaches zero beyond these limits. It is represented by the height of the broken line in Fig. 41·3 above the horizontal medial line (cf. **§ 9·3**). It will be seen from **(41·45)** that each particular value of A is propagated with the velocity

$$\Delta T/\Delta\Omega,$$

and this must of course be identical with the group velocity which may consequently be written

$$v = dv/dv', \quad . \quad . \quad . \quad . \quad . \quad \textbf{(41·46)}$$

since $\Delta T = 2\pi\Delta\nu$,

and $\Delta\Omega = 2\pi\Delta\nu'$,

where ν and ν' are respectively frequency and wave-number. It is easy to show that **(41·46)** is in agreement with **(41·44)**. We have in fact

$$dv/dv' = d(u\nu')/dv' = u + \nu' du/dv'$$

and ν' and dv' are respectively equal to $1/\lambda$ and $-d\lambda/\lambda^2$, so that on substitution we obtain **(41·44)**.

The simple group we have been describing is unlimited laterally, i.e. in directions perpendicular to the direction of propagation and when it is referred to rectangular axes of co-ordinates with the Z axis in the direction of propagation the amplitudes, a_s, are independent of x and y. If we wish to investigate the character of a group which is limited in all directions, i.e. in the X and Y directions as well as in the Z direction, we must extend the investigation of **§ 41·3** and study the Fourier expansion of a function like the $f(z)$ of **§ 41·3,** but described in the following way :

$$f(x, y, z) = 0 \begin{cases} -L < z < -l \\ l < z < L \\ -L' < x < -l' \\ l' < x < L' \\ -L'' < y < -l'' \\ l'' < y < L'' \end{cases}$$

$$f(x, y, z) = A \cos \Omega z \begin{cases} -l < z < l \\ -l' < x < l' \\ -l'' < y < l'' \end{cases}$$

The Fourier expansion for such a function may be written

$$f(x, y, z) = \sum_{p}\sum_{r}\sum_{s}\alpha_{p,r,s} \cos \frac{p\pi}{L'}x \cos \frac{r\pi}{L''}y \cos \frac{s\pi}{L}z. \quad (41{\cdot}47)$$

As before $\Omega_s(= s\pi/L)$ is very large, order + 1, but $\Omega_p(= p\pi/L')$ and $\Omega_r(= r\pi/L'')$ are only of the order O. The amplitudes associated with the values of p, r, and s of other orders than those mentioned are ignorable.

The symmetry of **(41·47)** enables us to put it in the form:

$$f(x, y, z) = \sum_{p}\sum_{r}\sum_{s}\alpha_{p,r,s} \cos \left(\frac{p\pi}{L'}x + \frac{r\pi}{L''}y + \frac{s\pi}{L}z\right), \quad (41{\cdot}48)$$

and we are thus led to a result analogous to **(41·43)** that a train or group of waves limited in all directions and describable within these limits by the expression

$$A \cos \Omega(z - ut),$$

A, Ω, and u being constants of the orders of magnitude 0, + 1, and 0 respectively, is practically identical with a superposition of infinitely extended simple plane waves, namely

$$\sum_{p}\sum_{r}\sum_{s}\alpha_{p,r,s} \cos \{\Omega_p x + \Omega_r y + \Omega_s(z - u_{p,r,s}t)\},$$

the Ω_p and Ω_r being of the order 0 while the Ω_s are as before of the order + 1. Hence we get

$$\psi = \sum_{p}\sum_{r}\sum_{s}\alpha_{p,r,s} \cos \omega_{p,r,s}\{\varepsilon_p x + \varepsilon_r y + \varepsilon_s(z - u_{p,r,s}t)\}, \quad (41{\cdot}49)$$

where

$$\omega_{p,r,s} = \sqrt{\Omega_p{}^2 + \Omega_r{}^2 + \Omega_s{}^2}\,|.$$

Clearly the ε_p and ε_r are small quantities of the order − 1, while the ε_s differ from unity by second order small quantities, i.e. by quantities of the order − 2.

This group is clearly a superposition of simple plane waves extending to infinity in all directions and travelling in directions which build up an infinitely narrow cone, since the cosine of the angle between any one of them and the direction of the Z axis differs from unity by a small number of the order − 2. It is on account of this last-named circumstance, which is due to the fact that Ω is very large, i.e. to the fact that λ is very small, that the group can only spread at an inappreciable rate and consequently behaves as if its boundary region were nearly rigid (cf. **§§ 27·1** and **28·2**).

§ 41·5. The Analogy between Classical Dynamics and Geometrical Optics

The important analogy which we are now going to study is more precisely described as an analogy between classical dynamics and a certain limiting case of wave propagation. As we have frequently pointed out, e.g. in **§§ 34·6, 39,** and **40·3,** we have been forced to depart in very important respects from the views of Young, Fresnel, and Maxwell about light, so that now we can hardly be said to regard light as an undulatory phenomenon in some kind of luminiferous medium (aether). This is not of course an ignoration of the fact that optical phenomena exhibit the familiar features associated with waves. Light waves are not physical entities in the same sense as are, for example, sound waves, but many of the laws governing optical phenomena are exactly those governing wave phenomena and thus many of the calculations in optics are effected by rules and formulae which are characteristic of waves, and it is only the undulatory aspect of optics which appears in the analogy we are about to describe and which was discovered by Hamilton early in the last century. Let us regard monochromatic light as constituted of such simple groups as we have described in **§ 41·4,** i.e. of groups of plane sinusoidal waves, limited in all directions and having a wave-length, λ, very small compared with the dimensions of the group. Each such group travels (in an isotropic medium) in a direction perpendicular to the plane wave fronts with the **group velocity,** $v = dv/dv'$, v being the **frequency** (number of vibrations per unit time) and v' being the **wave-number** (number of waves per unit length). In general, this velocity differs, as we have seen, from the **phase velocity** which is equal to v/v'. The two velocities happen to be equal in the absence of dispersion since then

$$du/dv' = 0,$$

and therefore

$$d(v/v')/dv' = 0,$$

or

$$dv/dv' - v/v' = 0.$$

Such a group is illustrated by G in Fig. 41·5(a) which is travelling from a place marked 1 to another place marked 2. The wave crests in the group are indicated by straight lines perpendicular to its direction of propagation. We have seen **(§ 41·4)** that it behaves as if its boundary were that of a rigid body. The special case of **geometrical optics** arises (cf. **§ 28·2**) when the simple groups, G, are so small compared with the dimensions of the lenses, apertures, etc., that they may be

regarded as points (Fig. 41·5(*b*)); or when the groups, *G*, if not very small, may be subdivided into smaller ones which are still simple groups though practically points in their spacial extension. Such minute groups are very like Newton's ' rays '. The difference in the meanings of the term ' ray ' as used by Newton, and as used at the present time, is just the difference between

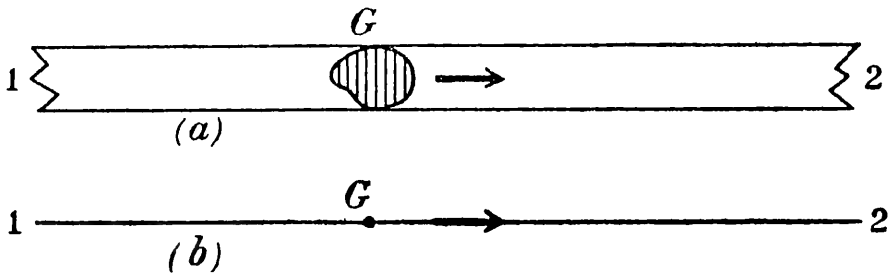

FIG. 41·5

the minute group, *G*, in Fig. 41·5(*b*) and the path along which it travels (cf. **§ 28·2**).

When the conditions just described, which define the scope of geometrical optics, are satisfied, interference and diffraction patterns are too small to be observed. Their dimensions are in fact of the order -1.

The laws of geometrical optics are contained in the single statement known as **Fermat's principle** (cf. **§§ 28·4** and **28·5**). We shall express it in the form

$$\delta\int_1^2 \nu'\, dl = 0, \qquad . \quad . \quad . \quad . \quad . \quad \textbf{(41·5)}$$

where ν' is the wave-number, i.e. the number of wave-lengths in the unit length, dl is an element of the path, Fig. 41·5(*b*), along which *G* travels, and the limits 1 and 2 represent two points on this path. Thus the integral $\int_1^2 \nu'\, dl$ represents the length of the actual path between 1 and 2, along which *G* travels, as measured by the number of waves in it. In the more usual language of geometrical optics it is the optical distance between 1 and 2 (cf. **§ 28·3**). The symbol δ denotes the difference between this length, i.e. the length of the **actual path,** and that of a **neighbouring path.** The actual path is defined or determined by certain parameters $\phi, \psi, \ldots$ and a neighbouring one by $\phi + \delta\phi$, $\phi + \delta\psi, \ldots$ where $\delta\phi, \delta\psi \ldots$ are of the order -1 (cf. **§ 28·4**). The statement **(41·5)** means, therefore, that the distances from 1 to 2 as measured along the actual path and along the neigh-

bouring one are equal, or, to be quite precise, differ by quantities of the order -2 at most.

The product $\nu'\,dl$ is an invariant and we may conveniently regard it as the scalar product $((\boldsymbol{\nu}', \mathbf{dl}))$ of two vectors $\boldsymbol{\nu}'$ and $\mathbf{dl}$ which, in an isotropic medium, have the same direction. If we use dq_α for the components of $\mathbf{dl}$—for example, dq_1, dq_2, and dq_3 mean, in a rectangular system of co-ordinates, dx, dy, and dz respectively—we may replace $\nu'\,dl$ by the sum

$$\nu_\alpha'\,dq_\alpha,$$

and so Fermat's principle becomes

$$\delta\int_1^2 \nu_\alpha'\,dq_\alpha = 0, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(41·51)}$$

where the repetition of the subscript α's serves to indicate summation. Fermat's principle, though often termed the principle of least time, in fact only determines the shape of the path between 1 and 2. Of course if the time were calculated on the assumption that G travels with the phase velocity, u, it would indeed have a stationary value and often a minimum value, for it would then be proportional to the number of waves, $\int_1^2 \nu_\alpha'\,dq_\alpha$, the sum of whose lengths is equal to that of the path. **But the time taken by G to traverse its path is determined by the group velocity** and Fermat's principle tells us nothing about this. The time is of no importance in the usual applications of geometrical optics; but here, where we are concerned with the nature of classical dynamics and the correspondence between it and geometrical optics, the group velocity, dv/dv' with which G travels along its path is of great importance. It is not, however, contained in Fermat's principle **(41·51)**.

The symbols δ and d will be better understood by reference to **§ 3·3.** The former refers to differences between neighbouring paths, the latter to differences between points on the same path, whether the actual one or a neighbouring one.

The statements **(41·5)** or **(41·51)** are associated with two highly important conditions which are usually tacitly understood. They are expressed by

$$\left.\begin{aligned} dv &= 0, \\ \delta v &= 0. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad \textbf{(41·511)}$$

Consequently ν_α', which is a function of ν and the q_α, may be regarded in **(41·51)** as a function of the q_α only, whenever

we are calculating the differences dv_α' or $\delta v_\alpha'$. The frequency, ν, is in fact a constant in the functional expressions for v_α' which we use in applying Fermat's principle.

The **principle of least action,** as it is usually termed **(8·636),** discovered by de Maupertius, can be expressed in the form :

$$\delta \int_1^2 p_\alpha \, dq_\alpha = 0, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(41·52)}$$

to which are attached the conditions

$$\left.\begin{aligned} dH &= 0, \\ \delta H &= 0, \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad \textbf{(41·521)}$$

where, of course, H means the energy of the system. Expressed in the form **(41·52)** with the conditions **(41·521)** it is strictly analogous to Fermat's principle. Like Fermat's principle, it merely gives the shapes of trajectories. It is true that velocities enter into it, since p_α means either *mass* × *velocity* or *moment of inertia* × *angular velocity* or the product of some other coefficient playing a part like mass, or moment of inertia, and the corresponding generalized velocity. But the velocities which thus enter do so in an adventitious manner, as will be better understood after Hamilton's principle has been discussed, in association with its geometrical optical parallisms. In order to bring out clearly the full scope of the analogy between classical dynamics and geometrical optics we must broaden the statement of Fermat's principle. The mere fact that we are dealing with waves suggests the following form :

$$\delta \int_1^2 (v_\alpha' \, dq_\alpha - \nu \, dt) = 0, \quad . \quad . \quad . \quad \textbf{(41·53)}$$

and this is strongly supported by the relativistic consideration that we should complete the 3-dimensional scalar product $v_\alpha' \, dq_\alpha$ by the addition of the fourth product, $-\nu \, dt$. If again we lay down the old conditions **(41·511)** it is clear that

$$\delta \int_1^2 v_\alpha' \, dq_\alpha$$

must vanish separately and consequently

$$\delta \int_1^2 \nu \, dt$$

must also vanish. It therefore follows from **(41·511)** that

$$\int_1^2 \delta\, dt = 0,$$

and therefore the two paths, the actual one and the neighbouring one, which are co-terminous in space, are also co-terminous in time. That is to say, they not only begin and end at common points, but start at the same time and end at the same time. But the extended principle (with the conditions **41·511**) includes no more of the physical features of the propagation than does the principle in its older and more restricted form. The extended form, like the older one, does no more than determine the shape of the path. It is different, however, when we alter the conditions attached to the application of **(41·53)**: as, for example, when we introduce the conditions

$$\left.\begin{aligned} &d\nu = 0, \\ &\delta\nu = \text{a small constant, order } -1, \end{aligned}\right\} \quad . \quad \textbf{(41·531)}$$

which are obviously compatible with one another. These new conditions include the old ones since it is open to us to adopt for $\delta\nu$ the value zero. The part, therefore, of the variation of **(41·53)** which is due to the δq_α must vanish separately as it did before; but when $\delta\nu$ is not zero we must have the extra contribution to the variation represented by

$$\delta\nu \int_1^2 \frac{\partial \nu_\alpha'}{\partial \nu}\, dq_\alpha - \delta\nu \int_1^2 dt$$

and, since this must vanish and $\delta\nu$ is arbitrary

$$\frac{dq_\alpha}{dt} = \frac{\partial \nu}{\partial \nu_\alpha'}. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(41·54)}$$

Our extended Fermat's principle therefore gives the correct expression for the velocity of the groups, G, of Fig. 41·5 (*b*). The exact correspondence of the equations **(41·54)** with Hamilton's canonical equations

$$\frac{dq_\alpha}{dt} = \frac{\partial H}{\partial p_\alpha}, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(41·541)}$$

should be observed (cf. **8·43**).

The condition $d\nu = 0$ **(41·531)** may be written

$$0 = \frac{\partial \nu}{\partial \nu_\alpha'} d\nu_\alpha' + \frac{\partial \nu}{\partial q_\alpha} dq_\alpha$$

and therefore by **(41·54)**

$$0 = \frac{dq_\alpha}{dt} dv_\alpha' + \frac{\partial v}{\partial q_\alpha} dq_\alpha,$$

and consequently

$$\frac{dv_\alpha'}{dt} = -\frac{\partial v}{\partial q_\alpha}. \quad . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(41·55)}$$

These equations correspond exactly to the canonical equations **(8·46)**, namely,

$$\frac{dp_\alpha}{dt} = -\frac{\partial H}{\partial q_\alpha}.$$

Finally, we may note that the statement **(41·53)** is in all respects exactly analogous to **Hamilton's principle,** which may be expressed in the form (cf. **§§ 8·6** and **35·7**)

$$\delta \int_1^2 (p_\alpha \, dq_\alpha - H \, dt) = 0. \quad . \quad . \quad . \quad \textbf{(41·56)}$$

It is worthwhile to observe that the velocities **(41·541)** can be deduced from Hamilton's principle **(41·56)**, just as the group velocities **(41·54)** can be deduced from the extended principle of Fermat **(41·53)**. In this respect Hamilton's principle is more comprehensive than that of Maupertius.

We may aptly describe the analogy or correspondence between classical mechanics and geometrical optics by saying that every purely mechanical problem—at any rate when it is associated with a conservative system—can be translated into a problem in geometrical optics, or into one which has exactly the same mathematical form as one in geometrical optics. The main features about this translation are the replacement of the generalized momenta, p_α, by wave-numbers, ν_α', and the energy, H, by a frequency, ν. In fact we may replace each momentum, p_α, by $h\nu_\alpha'$ and the energy by $h\nu$, where h is *any* constant.

§ 41·6. Wave-Mechanics

When the conditions described and illustrated diagrammatically in **§ 41·5** are not satisfied: namely, when λ is not a second order small quantity compared with the dimensions of the apparatus used, the lenses, apertures, etc., the laws of geometrical optics begin to fail. They still have an approximate validity when λ is not larger than a first order small quantity. They may then be supplemented by **Huygens' principle (31·85)**. The laws of geometrical optics and also Huygens' principle (cf. **§§ 31·7** and **31·8**) fail altogether when λ is of the same order

of magnitude as the apparatus. When this is the case, optical problems have to be attacked by suitable solutions of the partial differential equations (Maxwell's equations) governing them. Briefly, then, the comparatively simple laws of geometrical optics which, as we have seen, are analogous to those of classical mechanics, fail when the dimensions of the regions involved, the lengths of the optical paths, the apertures, etc., **are not sufficiently large** compared with the wave-length, λ, of the light.

Now we are familiar with a similar failure of classical dynamics, and of classical physical theory generally. This was first recognized and appreciated by Max Planck in his investigation of the problems of full radiation **(§ 39·9).**

Classical methods fail when applied to very small systems, atoms, electrons, etc. If now we bear in mind the remarkable correspondence between geometrical optics and Hamiltonian dynamics and also that both break down under similar circumstances we are naturally led to a new principle : **All dynamical problems can be translated into optical problems,** or more precisely into problems whose mathematical form is that of optical problems, i.e. of optical problems from the point of view of the undulatory theories of Young, Fresnel, or Maxwell. We have already anticipated this principle in **§ 9·4** and called it **Schroedinger's principle,** since his methods constitute virtually, if not actually, an application of it. When the dynamical problem happens, for example, to be a microphysical one the corresponding optical problem is **not one of geometrical optics but of undulatory optics.** This acceptance of a larger parallelism between mechanics and optics than that discovered by Hamilton constitutes the basis of **wave-mechanics.**

§ 41·7. de Broglie Waves

Louis de Broglie introduced his wave-mechanics in the attempt to reconcile the apparently contradictory undulatory and corpuscular features in light. He began by assuming that in the phenomena of light both waves and particles (photons) are associated. The energy in a beam of light he supposed, as did Einstein, to be seated in the photons, each having the energy $h\nu$, h being Planck's constant. He assigned to the waves the function of guiding the photons along lines normal to the wave fronts. Having arrived at this view of the nature of light and having reflected that, in our strong conviction, since the time of Young and Fresnel, of the undulatory nature of light, we have been blind to the possibility of the corpuscular elements in it, he suspected that we may have erred in the opposite sense in our

views of matter and thus have overlooked the possibility of waves being associated with elementary particles such as electrons. Just as in the case of photons he assigned the energy $h\nu$ to each elementary particle, ν being the frequency in the associated wave or simple group and h being Planck's constant. He identified its momentum with $h\nu'$, where ν' is the wave-number of the waves, and he supposed the velocity of the particle to be equal to the group velocity, $\partial\nu/\partial\nu'$, of the simple group. Finally, he gave the following relationship between group and phase velocity:

$$uv = c^2. \quad . \quad . \quad . \quad . \quad . \quad . \quad (41{\cdot}7)$$

This last formula is easily derived. The Hamiltonian correspondence gives

$$u = \nu/\nu' = h\nu/h\nu' = H/p.$$

Now we have seen that, for a particle,

$$H = mc^2.$$

Therefore $$u = mc^2/mv = c^2/v.$$

These brilliant predictions were completely confirmed by the experimental work of Davisson and Germer, G. P. Thomson, E. Rupp, and their successors.

The present view of the nature of the waves, often termed **de Broglie waves**, which are associated with elementary particles, differs from that originally held by de Broglie and also from that of Schroedinger (described later). We cannot now ascribe any physical reality to them any more than we can ascribe it to the waves in the geometrical optical picture of classical dynamics and they should be regarded as symbols used in the mathematical description of the characteristics of the phenomena with which we associate them.

§ 41·8. Photons and Elementary Particles

A parallel beam of light consists, according to the undulatory theory of optics (Young, Fresnel, Maxwell), of such simple groups as are described in **§ 41·5.** Let us consider what happens when it falls on a screen containing an aperture, e.g. a screen perpendicular to the direction in which the light is travelling and containing a rectangular aperture or slit. When the wave-length is very minute indeed (e.g. gamma rays), diffraction phenomena are not observed. The only effect of the aperture is to cut down the lateral dimensions of the beam. The photons which pass through the aperture simply travel straight onwards and their motion is governed by classical dynamics.[1]

[1] Classical dynamics includes relativistic dynamics.

On the other hand, with monochromatic light within the range of the visible spectrum in which the wave-length is much longer, it is easy to observe fringes **(§ 30·2)**, and of course even with such light the fringe system contracts as the aperture is made broader until it eventually becomes unobservable. Whether it can be observed or not is simply a question of the relative sizes of wave-length and breadth of aperture. The fringe system is not due of course to one group interfering with another. The only coherence **(§ 30·1)** is that between one part and another of the same group. Each group produces a complete fringe system. Further, there is only a single photon associated with each group. The distinctive features of wave-mechanics now begin to appear. The size of the wave-length, λ, has become significant. The phenomena which belong strictly to the realm of geometrical optics do not determine wave-lengths and we learned in **§ 41·5** that in the expression of the parallelism between geometrical optics and classical dynamics the constant, h, may have any value. It is different when we emerge from geometrical optics. We can now measure λ, and since in the parallelism, which we continue to accept as existing between optics and dynamics, $p = h\nu' = h/\lambda$, we find a precise value for h, since $h = p\lambda$. We thus learn that the parallelism between optics and mechanics gives a reason for the existence of this constant of Planck.

When the wave-length, λ, which we associate with a photon (or other elementary particle) is sufficiently small we may represent the particle by a simple group of practically punctual dimensions. The smallness of λ by comparison with the dimensions of the group will be assured when λ is a small quantity of the second order (i.e. of the order -2). The momentum of the photon is h/λ (or $h\nu'$) and therefore very great (macrophysical case). Its velocity and position may be identified with those of the simple group (G in Fig. 41·5(b)) and its motion is in accordance with classical dynamical principles. When λ is larger, for example, when λ is of the order 0, a simple group is only possible when it (the group) is large. The momentum of the photon is again h/λ, but very small (microphysical case). We may identify its velocity with that of the simple group, but its position has become vague. Indeed, when a beam of light passes through a slit there is nothing which determines the part of the fringe system (which, it will be remembered, is completely formed by the group representing a single photon) to which the photon travels. We do know, however, that the intensity of the light, for example, at places on a screen on which it is incident, is everywhere proportional to the square of the amplitude, A^2.

Therefore the number of photons falling on the unit area per unit time is proportional to A^2. We are consequently led to regard the value of A^2 at any point in a simple group representing a single photon or other elementary particle as a measure of the probability that the photon is there. More precisely

$$A^2\,dx\,dy\,dz \quad . \quad . \quad . \quad . \quad . \quad \textbf{(41·8)}$$

is the probability that the photon is in the volume element $dx\,dy\,dz$. Naturally we fix the absolute value of A^2 so that

$$\iiint A^2\,dx\,dy\,dz = 1$$

when the integration is extended over the whole of the region occupied by the wave-system which represents the photon and thus conforms with the convention which represents certainty by unity. We may do this because only relative values of A or A^2 are otherwise determined.

In this connexion it is of great interest to revert to formula **(41·34)** and multiply both sides by h. Thus

$$\Delta(h/\lambda) \times 2\Lambda \sim h$$

and 2Λ is the extreme length in the Z direction over which the simple group extends. It represents, in fact, the uncertainty, call it Δq, in the value of the Z co-ordinate of the photon (or elementary particle), and clearly $\Delta(h/\lambda)$ represents the uncertainty in the value of the corresponding momentum component. Hence

$$\Delta p \times \Delta q \sim h. \quad . \quad . \quad . \quad . \quad \textbf{(41·81)}$$

This means that, however accurately we may measure p, for example, i.e. however small the uncertainty, Δp, in its value may become, that of the conjugate quantity, namely, Δq, will become correspondingly greater. There is a lower limit, of the order of h, for their product. The result **(41·81)** is usually called Heisenberg's **uncertainty relation.** It is well illustrated by Heisenberg's 'Gedankenexperiment' with a gamma-ray microscope. The experiment consists in locating the position of a minute particle by means of an instrument of high resolving power—a microscope using light (photons) of very short wave-length. In Fig. 41·8 AB is the objective of the microscope and the particle which we are

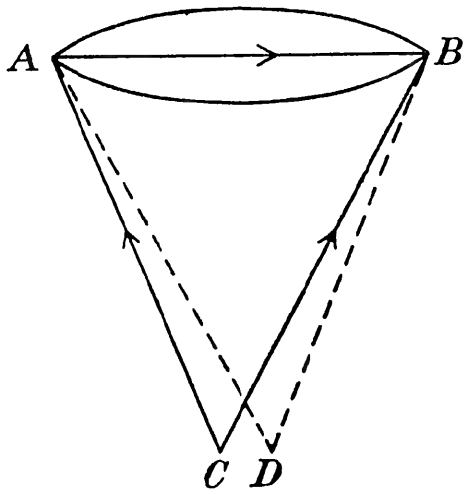

FIG. 41·8

locating is at C. The uncertainty, Δq, in the value of the co-ordinate of the particle parallel to AB may be represented by CD. Now

$$\frac{(CD)}{(AC)} \sim \frac{\lambda}{(AB)} \quad . \quad . \quad . \quad . \quad . \quad (41{\cdot}82)$$

(cf. **§ 30·5**). Suppose now that we happen to know very precisely the component of the momentum of the particle, parallel to AB, just before the observation of its position and that we succeed in observing the particle by means of the scattering into the objective of a single photon. The momentum of this photon will lie between that represented by the lines CA and CB. Its component parallel to CD is thus uncertain by an amount represented by AB. Therefore the component, parallel to CD, of the momentum of the particle has been modified, in consequence of the collision with the photon, by an amount lying between zero and AB. So that Δp is represented by AB. Now AC is a measure of the momentum of the scattered photon, or at any rate is of the same order of magnitude. This is not of course the momentum of the particle which is being located, so we shall call it p'. We have therefore from **(41·82)**

$$\frac{\Delta q}{p'} \sim \frac{\lambda'}{\Delta p}$$

or

$$\Delta p \times \Delta q \sim p'\lambda' \sim h,$$

in agreement with **(41·81).**

§ 41·9. An Illustration of the Wave Equation

In micromechanics a dynamical system is represented by a linear partial differential equation such as we have met in dealing with waves and our next problem is to discover a general method of constructing such an equation. As a guide we shall begin with a simple special case, that of a rigid body capable of rotating about a fixed axis and free from forces so that its angular momentum, p, and energy, H, are constants. The corresponding wave equation represents a propagation of sinusoidal waves with a phase velocity $v/v' = hv/hv' = H/p$ and may be written

$$\psi = A \cos \frac{2\pi}{h}(pq - Ht). \quad . \quad . \quad . \quad . \quad (41{\cdot}9)$$

This equation represents an extreme case, namely, that for which H and p are constant, and precisely given, and it extends from $q = -\infty$ to $q = +\infty$. It is what we have described as a simple wave. This case is so very simple that we have not

needed to set up a differential equation first in order to arrive at the result **(41·9)**; but in general, of course, we can only arrive at the result corresponding to **(41·9)** by first setting up and then solving a differential equation, just as in classical physical theory. Now the differential equation suggested for our study by **(41·9)** is obviously

$$\frac{\partial^2\psi}{\partial t^2} = u^2\frac{\partial^2\psi}{\partial q^2}, \quad . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(41·91)}$$

where u is a constant, namely H/p. The old way (cf. **§ 9**) of dealing with such an equation consists in finding particular solutions, ψ, each of which is a product of functions of one independent variable only. In the present example we search for a particular solution

$$\psi = T.Q,$$

where T is a function of t only and Q is a function of q only. When we substitute in the differential equation we get

$$Q\frac{d^2T}{dt^2} = u^2T\frac{d^2Q}{dq^2}$$

or, on dividing by TQ,

$$\frac{1}{T}\frac{d^2T}{dt^2} = \frac{u^2}{Q}\frac{d^2Q}{dq^2}.$$

This can be satisfied only by equating each side to the same arbitrary constant, thus

$$\frac{1}{T}\frac{d^2T}{dt^2} = k,$$

$$\frac{u^2}{Q}\frac{d^2Q}{dq^2} = k.$$

Hence

$$T = ae^{\sqrt{k}|t} + a'e^{-\sqrt{k}|t},$$

$$Q = be^{\frac{1}{u}\sqrt{k}|q} + b'e^{-\frac{1}{u}\sqrt{k}|t},$$

a, a', b, and b' being arbitrary constants. Thus we get, for example, the following particular solution of the differential equation:

$$\psi = TQ = Ae^{\sqrt{k}|\left(\frac{q}{u}-t\right)} \quad . \quad . \quad . \quad . \quad \textbf{(41·911)}$$

in which $A(=ba')$ is an arbitrary constant. Now the case of constant energy implies a frequency $\nu = H/h$, therefore

$$\sqrt{k}| = 2\pi i\nu = 2\pi iH/h. \quad . \quad . \quad . \quad \textbf{(41·92)}$$

When q increases by 2π the configuration of the physical system

we are dealing with, in this case a body rotating about a fixed axis, is exactly repeated so that Q must repeat its value by such a complete rotation. Therefore

$$be^{\frac{1}{u}\sqrt{k}|(q+2\pi)} = be^{\frac{1}{u}\sqrt{k}|q}$$

and consequently

$$e^{\frac{1}{u}\sqrt{k}|2\pi} = 1.$$

Hence it follows that

$$\frac{1}{u}\sqrt{k}\,|\,2\pi = 2\pi in$$

where n is any integer. Therefore

$$\sqrt{k}\,| = inu = inH/p. \quad . \quad . \quad . \quad . \quad \textbf{(41·921)}$$

On equating the two expressions, **(41·92)** and **(41·921)**, for $\sqrt{k}\,|$ we get

$$p = nh/2\pi. \quad . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(41·93)}$$

This is of course precisely what the older quantum theory gave for the momentum in such a case.

The relation between p and H in the case of a rigid body rotating about a fixed axis is

$$H = p^2/2I,$$

where I is the moment of inertia and a constant. Hence by **(41·93)**

$$H = n^2h^2/8\pi^2I \quad . \quad . \quad . \quad . \quad . \quad \textbf{(41·94)}$$

(cf. **§§ 40·2, 40·8,** and **40·9**).

The old quantum theory obtained these results by prescribing for each phase integral $\oint p_\alpha\, dq_\alpha$ the value $n_\alpha h$ where n_α is a positive integer or zero. In the present example there is only one phase integral and the angular momentum, p, is constant. Therefore

$$\oint p\, dq = p\oint dq = p2\pi = nh.$$

In the more general case of a free axis we shall see (cf. **42·3**) that, instead of **(41·94)**, we get for the energy

$$H = n(n+1)h^2/8\pi^2I. \quad . \quad . \quad . \quad . \quad \textbf{(41·95)}$$

When we substitute the value **(41·92)** for $\sqrt{k}\,|$ in the particular solution **(41·911)** we get

$$\psi = Ae^{\frac{2\pi i}{h}(pq-Ht)}. \quad . \quad . \quad . \quad . \quad \textbf{(41·96)}$$

This particular solution is complex and we may compare it with

the complex solutions of the differential equations we met with in the study of electromagnetic waves and oscillations. The advantages of such solutions are even more conspicuous in wave-mechanics.

When we differentiate ψ in **(41·96)** partially with respect to q and t we obtain

$$\frac{\partial \psi}{\partial q} = \frac{2\pi i}{h} p\psi,$$

and

$$\frac{\partial \psi}{\partial t} = -\frac{2\pi i}{h} H\psi,$$

so that we may regard p and H as equivalent to the operations

$$\left.\begin{aligned} p &\equiv (h/2\pi i)\frac{\partial}{\partial q} \\ \text{and} \qquad H &\equiv -(h/2\pi i)\frac{\partial}{\partial t} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad \textbf{(41·97)}$$

respectively carried out on the function ψ of **(41·96)**.

The particular solution **(41·911)** of the differential equation **(41·91)** is a solution whatever value may be assigned to $\sqrt{k}$; but, as we have just seen, the conditions associated with the special problem we are dealing with restrict the admissible values of this constant to an integral multiple of iu or iH/p **(41·921).** The most general admissible solution will of course be a sum

$$\psi = \sum_{\alpha} c_\alpha T_\alpha Q_\alpha. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(41·98)}$$

Such functions as T_α and Q_α are known as **proper functions** and the associated special value of k, or $\sqrt{k}$, is called a **proper value.** The corresponding German terms are *Eigenfunktion* and *Eigenwert.*

CHAPTER XI

OUTLINE OF WAVE-MECHANICS

§ 42. The Wave Equation

WE cannot of course deduce the mode of setting up the general differential equation, which will fit all cases, from the study of one or more particular cases like that in the last section. But such particular cases provide us with helpful suggestions which we can usefully follow. The most obvious of these is the following generalization of **(41·96)**:

$$\psi = A(q_1,\ q_2,\ \ldots\ t)e^{\frac{2\pi i}{h}S} \quad . \quad . \quad . \quad . \quad \textbf{(42)}$$

in which A and S are real functions of the real variables $q_1,\ q_2\ \ldots\ t$ and S is the Hamiltonian principal function

$$S = \int (p_1\,dq_1 + p_2\,dq_2\ \ldots\ - H\,dt)$$

(cf. **§§ 8·6** and **8·7**). For the present let us suppose the system is a particle of mass m and that q_1, q_2, and q_3 are the rectangular co-ordinates of its position (more usually represented by x, y, and z). We shall also assume that the particle is in a field of force with which a potential energy function, V, is associated. According to the classical (non-relativistic) theory the behaviour of the particle is subject to Hamilton's equation

$$\frac{1}{2m}\sum_{\alpha}\left(\frac{\partial S}{\partial q_\alpha}\right)^2 + V + \frac{\partial S}{\partial t} = 0, \quad . \quad . \quad . \quad \textbf{(42·01)}$$

(cf. **8·68**). We shall now carry out differential operations on **(42)** and build up a differential equation under the guidance of the principle that the classical theory must hold in the macrophysical case, i.e. **(42·01)** must emerge when m is very large. We easily find that

$$\frac{\partial^2\psi}{\partial {q_1}^2} = \left[\left\{\frac{1}{A}\frac{\partial^2 A}{\partial {q_1}^2} - \frac{4\pi^2}{h^2}\left(\frac{\partial S}{\partial q_1}\right)^2\right\} + i\left\{\frac{4\pi}{Ah}\frac{\partial A}{\partial q_1}\frac{\partial S}{\partial q_1} + \frac{2\pi}{h}\frac{\partial^2 S}{\partial {q_1}^2}\right\}\right]\psi$$

and

$$\frac{\partial\psi}{\partial t} = \left\{\frac{1}{A}\frac{\partial A}{\partial t} + \frac{2\pi i}{h}\frac{\partial S}{\partial t}\right\}\psi.$$

Hence

$$-\frac{h^2}{8\pi^2 m}\nabla^2\psi + V\psi + \frac{h}{2\pi i}\frac{\partial\psi}{\partial t} \equiv$$

$$\left[\left\{\frac{1}{2m}\sum_\alpha\left(\frac{\partial S}{\partial q_\alpha}\right)^2 + \left(V - \frac{h^2}{8\pi^2 m}\frac{1}{A}\nabla^2 A\right) + \frac{\partial S}{\partial t}\right\}\right.$$

$$\left. - i\left\{\frac{h}{2\pi A m}\sum_\alpha \frac{\partial A}{\partial q_\alpha}\frac{\partial S}{\partial q_\alpha} + \frac{h}{4\pi m}\nabla^2 S + \frac{h}{2\pi A}\frac{\partial A}{\partial t}\right\}\right]\psi$$

$$\equiv (M - iN)\psi. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(42·02)}$$

It is now clear that if we adopt the equation

$$-\frac{h^2}{8\pi^2 m}\nabla^2\psi + V.\psi + \frac{h}{2\pi i}\frac{\partial\psi}{\partial t} = 0,$$

or

$$\boxed{\nabla^2\psi - \frac{8\pi^2 m}{h^2}V\psi + \frac{4\pi m i}{h}\frac{\partial\psi}{\partial t} = 0,} \quad . \quad . \quad \textbf{(42·03)}$$

we shall satisfy the requirement that Hamiltonian dynamics will hold in the macrophysical case (m very large). since M and N must vanish separately and M reduces, when m is large (compared with h), to **(42·01).**

It should be noted here that **(42·03)** is precisely what we get when we regard the left-hand side of **(42·01)** as an operator and replace each p_α, or $\partial S/\partial q_\alpha$, by the operator **(41·97),** and when we similarly replace $\partial S/\partial t$, which is equal to $-H$, by the appropriate operator **(41·97).**

Since the right-hand side of **(42·02),** which we have written $M - iN$ vanishes, it is clear that N must vanish separately. Thus

$$\frac{h}{2\pi A m}\sum_\alpha \frac{\partial A}{\partial q_\alpha}\frac{\partial S}{\partial q_\alpha} + \frac{h}{4\pi m}\nabla^2 S + \frac{h}{2\pi A}\frac{\partial A}{\partial t} = 0,$$

or with a very little reduction

$$\Sigma p_\alpha \frac{\partial(A^2)}{\partial q_\alpha} + A^2\Sigma\frac{\partial p_\alpha}{\partial q_\alpha} + m\frac{\partial(A^2)}{\partial t} = 0.$$

If we revert to the more usual notation for rectangular co-ordinates (x, y, z instead of q_1, q_2, q_3) we may write the equation in the form

$$m\left\{v_x\frac{\partial(A^2)}{\partial x} + v_y\frac{\partial(A^2)}{\partial y} + v_z\frac{\partial(A^2)}{\partial z}\right\}$$

$$+ mA^2\left\{\frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z}\right\} + m\frac{\partial(A^2)}{\partial t} = 0$$

and finally

$$\boxed{\mathbf{div}\,(A^2\mathbf{v}) + \frac{\partial(A^2)}{\partial t} = 0.} \quad . \quad . \quad . \quad \textbf{(42·04)}$$

This is a very significant result. It asserts that a certain quantity whose density is A^2 and whose velocity is $\mathbf{v}$ is conserved. It justifies our use of $A^2\,dx\,dy\,dz$ to express the probability that the particle is situated within $dx\,dy\,dz$ because it assures us that if $\iiint A^2\,dx\,dy\,dz = 1$ at any instant, when the integration is extended over the whole region where A^2 differs from zero, it must also always be equal to 1.

§ 42·1. Schroedinger's Amplitude Equation

When we take proper solutions of **(42·03)** for which the time factor is the exponential $e^{-2\pi i\nu t}$, so that $\psi = ae^{-2\pi i\nu t}$, we may replace $\partial\psi/\partial t$ in **(42·03)** by $-2\pi i\nu\psi$.

We thus obtain on substituting in **(42·03)**

$$\nabla^2\psi + \frac{8\pi^2 m}{h^2}(H - V)\psi = 0$$

or since $\psi = ae^{-2\pi i\nu t}$

$$\boxed{\nabla^2 a + \frac{8\pi^2 m}{h^2}(H - V)a = 0.} \quad . \quad . \quad . \quad \textbf{(42·1)}$$

This is **Schroedinger's amplitude equation.**

§ 42·2. The Simple Harmonic Oscillator

It is appropriate to illustrate Schroedinger's theory by applying it, in the first instance, to the case of simple harmonic motion. It will be remembered **(§ 39·9)** that Planck assumed the energy of a simple harmonic system to be $nh\nu_0$ where ν_0 is its frequency of oscillation and n is a positive integer, including zero.[1] Schroedinger's equation **(42·1)** for the present case is

$$\frac{d^2a}{dq^2} + \frac{8\pi^2 m}{h^2}\left(E - \frac{\mu}{2}q^2\right)a = 0, \quad . \quad . \quad . \quad \textbf{(42·2)}$$

where a is the amplitude in the expression

$$\psi = ae^{-2\pi i\nu t}, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(42·21)}$$

[1] Planck modified this in his later work to $(n + \frac{1}{2})h\nu_0$, thus anticipating, as we shall see, the pronouncement of wave-mechanics.

ν is the frequency of the associated de Broglie wave and $h\nu$ the energy of the system corresponding to the proper solution **(42·21).**

It is convenient to replace the variable, q, by a new one, x, so chosen as to simplify **(42·2)** to

$$\frac{d^2a}{dx^2} + (R - x^2)a = 0. \quad . \quad . \quad . \quad . \quad \textbf{(42·22)}$$

We can do this by the substitution

$$x = \beta q,$$

if we give to β a suitable constant value. On substituting we get

$$\frac{d^2a}{dx^2} + \left(\frac{8\pi^2 mE}{\beta^2 h^2} - \frac{4\pi^2 m\mu}{\beta^4 h^2}x^2\right)a = 0.$$

We must clearly assign to β such a value that

$$4\pi^2 m\mu/\beta^4 h^2 = 1.$$

We thus get

$$\frac{d^2a}{dx^2} + \left(4\pi\sqrt{\frac{m}{\mu}}\frac{E}{h} - x^2\right)a = 0,$$

or, since

$$4\pi\sqrt{m/\mu} = 2/\nu_0,$$

where ν_0 is the frequency of the oscillator,

$$\frac{d^2a}{dx^2} + \left(\frac{2E}{h\nu_0} - x^2\right)a = 0. \quad . \quad . \quad . \quad \textbf{(42·23)}$$

The amplitude, a, in this equation is illustrated by Q in **§ 41·9.** Just as in the case of Q the amplitude a has to satisfy certain conditions, namely, it has to be one-valued, finite and continuous for all real values of q between $-\infty$ and $+\infty$, i.e. for all real values of x between these limits, and it must approach the limit zero when q (or x) approaches ∞. Indeed, the integral $\int_0^\infty a^2\,dq$ has to be finite so that we may make it unity by a suitable choice of the otherwise undetermined constant with which any solution may be multiplied.

It is easy to see that

$$a = e^{-\frac{x^2}{2}}$$

will satisfy **(42·23)** when

$$2E/h\nu_0 = 1,$$

so that a possible energy value is

$$E = h\nu_0/2.$$

To deal more generally with the problem we shall try to get solutions of the form

$$a = ve^{-\frac{x^2}{2}},$$

where v is a (so far) unknown function of x. When we substitute this expression for a in **(42·23)** we obtain

$$e^{-\frac{x^2}{2}}\left(\frac{d^2v}{dx^2} - 2x\frac{dv}{dx} + [R-1]v\right) = 0,$$

where $R = 2E/h\nu_0$. Hence we find for v the differential equation

$$\frac{d^2v}{dx^2} - 2x\frac{dv}{dx} + [R-1]v = 0. \quad . \quad . \quad . \quad \textbf{(42·24)}$$

Following a well-known method, we expand v in the form:

$$v = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3 + \ldots,$$

the coefficients, α, being constants which we have to determine. We then substitute this expression for v in the differential coefficients dv/dx and d^2v/dx^2 and substitute in **(42·24)**. The equation is satisfied by equating the coefficient of each resulting power of x to zero. The coefficient of x^n, for example, is

$$(R - 1 - 2n)\alpha_n + (n+1)(n+2)\alpha_{n+2}. \quad . \quad \textbf{(42·25)}$$

This must be equated to zero.

Thus there are in the expression for v two coefficients which, if we are only concerned in finding a solution of the differential equation **(42·24)**, may be given any values we choose. We might for example give α_0 any value we choose, in which case all the α_s, where s is even, are determined. Similarly the α_s where s is odd might be fixed by assigning an arbitrary value to α_1. Instead of proceeding in this way, suppose we assign the value zero to α_1, in which case all the α_s for which s is odd vanish, and the value zero to α_{n+2} where n is any even integer we care to choose. Therefore by **(42·25)** we have

$$(R - 1 - 2n)\alpha_n = 0.$$

This equation will be satisfied even when α_n itself is not zero provided

$$R = 2n + 1 \quad . \quad . \quad . \quad . \quad . \quad \textbf{(42·251)}$$

or

$$R - 1 = 2n.$$

Equation **(42·25)** gives us further for this value of $R - 1$

$$\{2n - 2(n-2)\}\alpha_{n-2} + (n-1)(n)\alpha_n = 0,$$

or

$$4\alpha_{n-2} + n(n-1)\alpha_n = 0,$$

thus α_{n-2} is expressed in terms of α_n. And so we may determine all the α_s for which s is an even number less than n. Further

$$\{2n - 2(n+2)\}\alpha_{n+2} + (n+3)(n+4)\alpha_{n+4} = 0.$$

Since α_{n+2} is zero, so also is α_{n+4} and therefore all the coefficients α_s where s is even and greater than n. Thus we get for v a series with a finite number of terms in it, $\alpha_n v^n$ being the last one. Therefore the solution

$$a = e^{-\frac{x^2}{2}} v$$

satisfies the requirement of being finite, continuous and only valued for all values of x between zero and infinity. We reach a similar result, of course, by proceeding in such a way that all the even α_s are equal to zero and choosing R so that n in **(42·251)** is odd.

Equation **(42·251)** gives us

$$E = (n + \tfrac{1}{2})h\nu_0. \quad . \quad . \quad . \quad . \quad \textbf{(42·26)}$$

Only when the energy, E, has one of the values **(42·26)** and for no other values of E do the particular (proper) solutions, a, of the amplitude equation satisfy the condition of being finite and continuous for all values of q and of approaching the limit zero when $q \longrightarrow \infty$.

The result **(42·26)** differs slightly from that obtained by applying the old quantum rule $\oint p dq = nh$, namely, $E = nh\nu_0$.

When the oscillator is emitting radiation, i.e. photons, the energy of each photon is equal to $h\nu$, where ν is the frequency of the radiation, we have

$$h\nu = (n_2 + \tfrac{1}{2})h\nu_0 - (n_1 + \tfrac{1}{2})h\nu_0$$

or

$$\nu = (n_2 - n_1)\nu_0.$$

The only permissible transitions, as the correspondence principle has already indicated, are those for which $n_2 - n_1$ is unity, so that

$$\nu = \nu_0.$$

How this emerges from wave-mechanics will be explained in **§ 42·8.**

The functions, v, which have appeared in this problem of the oscillator are known as **Hermitian functions.**

§ 42·3. The Rotator

Let us now study the case of a particle which is forced by constraints to keep to the surface of a sphere whose centre is fixed and suppose the particle subject to no forces except those due to the constraints. We may equate its potential energy to zero and consequently its Schroedinger equation becomes

$$\nabla^2 a + \frac{8\pi^2 m}{h^2} E a = 0 \quad . \quad . \quad . \quad . \quad \textbf{(42·3)}$$

It is now convenient to express the Laplacian $\nabla^2 a$, which in rectangular co-ordinates is $\frac{\partial^2 a}{\partial x^2} + \frac{\partial^2 a}{\partial y^2} + \frac{\partial^2 a}{\partial z^2}$, in terms of polar co-ordinates. This can most easily be done by applying Gauss's Theorem **(3·01)** to the product $\nabla^2 a\, r^2\, dr \sin\theta\, d\theta\, d\phi$. This product is equal to the surface integral $\frac{\partial a}{\partial n} dS$ integrated over the faces of the volume element, n representing the outward normal. The surfaces of the element are :

$$(a) \qquad r\, d\theta \times r \sin\theta\, d\phi$$

and a corresponding opposite one separated from it by the distance dr ;

$$(b) \qquad r \sin\theta\, d\phi \times dr$$

and opposite it one separated by the distance $r\, d\theta$, and, lastly,

$$(c) \qquad dr \times r\, d\theta,$$

and opposite this a similar surface separated from it by the distance $r \sin\theta\, d\phi$.

Applying Gauss's theorem, we get from (a)

$$r\, d\theta \times r \sin\theta\, d\phi \left\{ - \frac{\partial a}{\partial r} \right\}$$

or

$$- r^2 \frac{\partial a}{\partial r} \sin\theta\, d\theta\, d\phi$$

and from the opposite face :

$$\left\{ r^2 \frac{\partial a}{dr} + \frac{\partial}{\partial r}\left(r^2 \frac{\partial a}{\partial r} \right) dr \right\} \sin\theta\, d\theta\, d\phi.$$

Altogether from (a) and its opposite face

$$\left\{ \frac{\partial^2 a}{\partial r^2} + \frac{2}{r} \frac{\partial a}{\partial r} \right\} r^2\, dr \sin\theta\, d\theta\, d\phi.$$

If we divide by the volume element, $r^2\, dr \sin\theta\, d\theta\, d\phi$, we get the corresponding contribution to the Laplacian, namely,

$$\frac{\partial^2 a}{\partial r^2} + \frac{2}{r} \frac{\partial a}{\partial r}. \qquad \ldots\ldots \quad \textbf{(42·31)}$$

Similarly, from the face (b) and its opposite face we get

$$\frac{1}{r^2 \sin\theta} \frac{\partial}{\partial\theta}\left(\sin\theta \frac{\partial a}{\partial\theta} \right). \qquad \ldots \quad \textbf{(42·311)}$$

Lastly, from (*c*) and its opposite face we similarly find as the contribution to the Laplacian

$$\frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 a}{\partial \phi^2}. \quad . \quad . \quad . \quad . \quad . \quad (42{\cdot}312)$$

The complete Laplacian is therefore, in these co-ordinates,

$$\frac{\partial^2 a}{\partial r^2} + \frac{2}{r}\frac{\partial a}{\partial r} + \frac{1}{r^2 \sin \theta}\frac{\partial}{\partial \theta}\left(\sin \theta \frac{\partial a}{\partial \theta}\right) + \frac{1}{r^2 \sin^2 \theta}\frac{\partial^2 a}{\partial \phi^2}. \quad . \quad (42{\cdot}32)$$

In the case we are now studying r is constant and **(42·3)** becomes

$$\frac{\partial^2 a}{\partial \theta^2} + \frac{\cos \theta}{\sin \theta}\frac{\partial a}{\partial \theta} + \frac{1}{\sin^2 \theta}\frac{\partial^2 a}{\partial \phi^2} + \frac{8\pi^2 m r^2}{h^2}.E.a = 0,$$

or

$$\frac{\partial^2 a}{\partial \theta^2} + \frac{\cos \theta}{\sin \theta}\frac{\partial a}{\partial \theta} + \frac{1}{\sin^2 \theta}\frac{\partial^2 a}{\partial \phi^2} + Ma = 0, \quad . \quad (42{\cdot}33)$$

where

$$M \equiv \frac{8\pi^2 I}{h^2}.E \quad . \quad . \quad . \quad . \quad (42{\cdot}331)$$

and I is the constant (moment of inertia) mr^2.

Following the usual method, we shall endeavour to find a particular solution

$$a = \Theta.\Phi$$

where Θ depends on θ only and Φ on ϕ only. On substituting in **(42·33)** and dividing through by the product $\Theta.\Phi$ we get

$$\frac{1}{\Theta}\frac{d^2\Theta}{d\theta^2} + \frac{\cos \theta}{\Theta \sin \theta}\frac{d\Theta}{d\theta} + \frac{1}{\sin^2 \theta}\frac{1}{\Phi}\frac{d^2\Phi}{d\phi^2} + M = 0,$$

or

$$\frac{\sin^2 \theta}{\Theta}\frac{d^2\Theta}{d\theta^2} + \frac{\cos \theta \sin \theta}{\Theta}\frac{d\Theta}{d\theta} + \frac{1}{\Phi}\frac{d^2\Phi}{d\phi^2} + M \sin^2 \theta = 0.$$

We can satisfy this equation by

$$\frac{1}{\Phi}\frac{d^2\Phi}{d\phi^2} = -m^2 . \quad . \quad . \quad . \quad . \quad (42{\cdot}34)$$

and

$$\frac{\sin^2 \theta}{\Theta}\frac{d^2\Theta}{d\theta^2} + \frac{\cos \theta \sin \theta}{\Theta}\frac{d\Theta}{d\theta} + M \sin^2 \theta = m^2, \quad (42{\cdot}35)$$

where m^2 is any constant.

We require of a $(= \Theta\Phi)$ that it shall be finite and continuous for all values of θ and ϕ and, moreover, one-valued for all directions. We must consequently make m^2 a positive integer, so that

$$\Phi = B \cos (m\phi - \eta),$$

where B and η are constants. The function Φ will thus repeat its values when ϕ is changed by 2π.

Turning to **(42·35)**, we have

$$\frac{d^2\Theta}{d\theta^2} + \frac{\cos\theta}{\sin\theta}\frac{d\Theta}{d\theta} + \left(M - \frac{m^2}{\sin^2\theta}\right)\Theta = 0, \qquad \textbf{(42·351)}$$

and, when we represent $\cos\theta$ by x and substitute in this equation,

$$(1 - x^2)\frac{d^2\Theta}{dx^2} - 2x\frac{d\Theta}{dx} + \left(M - \frac{m^2}{1 - x^2}\right)\Theta = 0. \qquad \textbf{(42·36)}$$

This equation is dealt with in great detail in treatises on spherical harmonics. We can best deal with it by writing

$$\Theta = (1 - x^2)^{\frac{m}{2}}v$$

and substituting in **(42·36)**. When we do this we get for the function v, the equation

$$(1 - x^2)\frac{d^2v}{dx^2} - 2(m + 1)x\frac{dv}{dx} - (M - m(m + 1))v = 0. \qquad \textbf{(42·37)}$$

We can find a solution of it by representing v as a power series in x, just as we did in the case of equation **(42·24)**. Thus we write

$$v = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3 + \ldots, \qquad \textbf{(42·371)}$$

substitute this for v in **(42·37)** and equate the coefficients of each power of x in the resulting expression to zero. The coefficient of x^n is

$$\{M - m(m + 1) - 2n(m + 1) - n(n - 1)\}\alpha_n + (n + 1)(n + 2)\alpha_{n+2}. \qquad \textbf{(42·38)}$$

This and all other such coefficients must vanish. Therefore all the coefficients of the even powers of x in the expression for v can be expressed in terms of any one of them which may be chosen arbitrarily. Similarly all the coefficients of the odd powers of x in the expression for v can be expressed in terms of any one of them. The expression for v is therefore the sum of two parts: one containing even powers of x and an arbitrary constant, one of the even α's, and the other containing only odd powers of x and one of the odd α's.

Since the independent variable θ is confined to the range of real values between 0 and π, the variable x ranges over the real values between $+1$ and -1 only. Within these limits the function v must be finite, continuous and one-valued. We can secure this in the following way: let us take α_0 to be zero, in which case all the even coefficients must be zero, since **(42·38)** must be zero. We may then, for example, choose that any one of the odd coefficients, say α_{n+2}, is zero. The series for v then

consists of a finite number of odd powers of x ending with $\alpha_n x^n$ (n supposed to be odd). Equating **(42·38)** to zero, we get

$$M = (m + n)(m + n + 1). \quad . \quad . \quad \textbf{(42·381)}$$

Similarly we may, alternatively, choose α_1 to be zero, with the consequence that all the odd coefficients in the expression for v vanish. If now we choose that α_{n+2} (n being any even number) shall be zero, the resulting series for v again consists of a finite number of terms ending with $\alpha_n x^n$ (n being even), and once again we get **(42·381)**. We may just as well write **(42·381)** in the form

$$M = l(l + 1). \quad . \quad . \quad . \quad . \quad \textbf{(42·382)}$$

where l is an integer.

Thus by **(42·331)**

$$\frac{8\pi^2 \mathrm{I}}{h^2} E = l(l + 1)$$

or

$$E = \frac{l(l + 1)h^2}{8\pi^2 I}. \quad . \quad . \quad . \quad . \quad \textbf{(42·39)}$$

This is the result anticipated in **§ 41·9.**

§ 42·4. Band Spectra

The simplest type of band spectrum receives an immediate explanation by this last result. We may regard **(42·39)** as expressing the energy of rotation of the molecules in a gas or vapour. The integer, l, as we shall see in **§ 42·8,** can only change by unity. Thus when a rotating molecule emits radiation (or photons) the energy of one of these is

$$h\nu = \{l(l + 1) - (l - 1)l\}h^2/8\pi^2 I,$$

if we assume that the moment of inertia, I, of the molecule does not change when such a transition occurs. Hence the possible frequencies of the emitted radiation are

$$\nu = lh/4\pi^2 I. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(42·4)}$$

Thus such a spectrum consists of equally spaced lines (as measured by the frequency intervals). Such purely rotational spectra have been observed both as emission and as absorption spectra of various gases and vapours in the extreme infra-red.

If we suppose the molecule to be capable of simple harmonic vibrations, as well as of rotation, we get an explanation of another type of band spectrum. The expression for the energy of the molecule now becomes

$$(n + \tfrac{1}{2})h\nu_0 + l(l + 1)h^2/8\pi^2 I.$$

Both n and l can only change by unity **(§ 42·8).** When ν_0 is big enough the integer l may *increase* by unity in the emission of radiation as well as decrease. Thus we have for the frequency, ν, of the emitted radiation

$$h\nu = \{(n + \tfrac{3}{2}) - (n + \tfrac{1}{2})\}h\nu_0 + \{(l + 1)(l + 2) - l(l + 1)\}\frac{h^2}{8\pi^2 I},$$

and also

$$h\nu = \{(n + \tfrac{3}{2}) - (n + \tfrac{1}{2})\}h\nu_0 + \{(l - 1)l - l(l + 1)\}\frac{h^2}{8\pi^2 I},$$

or

$$\left.\begin{aligned} &\nu = \nu_0 + (l + 1)h/4\pi^2 I \\ \text{and} \qquad &\\ &\nu = \nu_0 - lh/4\pi^2 I. \end{aligned}\right\} \quad . \quad . \quad . \quad \textbf{(42·41)}$$

The spectrum therefore consists of two branches, one corresponding to each of the expressions **(42·41).** The individual lines (which correspond to different values of the integer, l) are separated by the common frequency difference $h/4\pi^2 I$. The formulae **(42·41)** indicate apparently a central frequency, ν_0. The observations, however, while they confirm **(42·41)** in other respects, indicate that this frequency is missing. We can account for this if we suppose that negative values of l are not permissible. Thus in the upper formula, where l represents the final value of the integer, $\nu = \nu_0$ would mean $l = -1$, while in the second formula, since l would have to be zero, the value of the integer before the transition would have to be -1. The reason for the absence of negative values of l is given in **§ 42·5.**

The relative intensities of the lines in bands like **(42·41)** can be explained at least approximately with the aid of Maxwell's distribution law **(12·16).** The intensities rise on either side of the position of the missing line, ν_0, reach a maximum and then diminish.

In addition to the two types of bands described above, still another type is well known, and indeed more familiar, since it is observable in the visible region. The classical example of this type is furnished by the cyanogen bands. Such bands are called electronic bands since their production is believed to be associated with electron transitions within the emitting molecule and a *consequent change in its moment of inertia.* The frequencies in these bands may be represented by

$$\nu = \nu_0 + \nu_e + [\textbf{rotational terms}]. \quad . \quad . \quad \textbf{(42·42)}$$

The frequency ν_0 is the vibrational one of **(42·41),** while ν_e is

the contribution due to the electron transition. The rotational terms arise out of

$$\frac{l'(l'+1)h}{8\pi^2 I'} - \frac{l(l+1)h}{8\pi^2 I},$$

which is the difference of the initial and final rotational energy after it has been divided by h, to give the contribution to the corresponding frequency.

There are three cases:

$$\begin{array}{lll} (1) & l' = l+1, & (R) \\ (2) & l' = l-1, & (P) \\ (3) & l' = l, & (Q) \end{array} \quad . \quad . \quad . \quad . \quad \textbf{(42·43)}$$

They are referred to as the R, P, and Q branches respectively. They may be expressed as follows:

$$\left.\begin{array}{l} (R)\ \nu = \nu_0 + \nu_e + a(l+1) + b(l+1)^2, \\ (P)\ \nu = \nu_0 + \nu_e - al + bl^2, \\ (Q)\ \nu = \nu_0 + \nu_e + bl + bl^2, \end{array}\right\} \quad . \quad . \quad \textbf{(42·44)}$$

where

$$a = \frac{h}{8\pi^2}\left(\frac{1}{I'} - \frac{1}{I}\right)$$

$$b = \frac{h}{8\pi^2}\left(\frac{1}{I'} + \frac{1}{I}\right)$$

and I' and I are the molecular moments of inertia before and after the transition respectively.

The vibrational frequencies are not necessarily simple harmonic. Their Fourier resolution, however, represents them as a superposition of simple harmonic vibrations whose frequencies are integral multiples of a fundamental one. There is consequently a band, the frequencies of whose components are given by **(42·44)**; for every one of these vibrational frequencies. They form, collectively, a **band system.**

When the integral values of l are plotted against the corresponding values of ν, as calculated by **(42·44)**, each branch, R, P, and Q, is represented by a parabola. The values of l are subject to restrictions similar to those associated with **(42·41)**. The branch, R, with the permitted values of $l+1$ includes the apex of this parabola and thus explains one of the most characteristic features of these bands. The lines are crowded together at the apex of the parabola, R, thus producing the **head** of the band.

The relative intensities of the lines vary from one value of l to another in a similar way to that mentioned in connexion with

vibration-rotation bands **(42·41)** and the line for which $l = 0$ is also missing.

The vibration-rotation bands of HCl (in the infra-red) show each line accompanied by a neighbouring weaker one of lower frequency. This is accounted for by two separate values of the vibration frequency ν_0 in consequence of the fact that there are two **isotopes** of Cl, namely, Cl_{35} and Cl_{37}. The former, having a smaller mass, has a correspondingly higher ν_0. The relative numbers of these isotopes is $Cl_{35} : Cl_{37} = 3 : 1$, and the corresponding intensities are in accord with this ratio.

§ 42·5. Kepler Motion and the Hydrogen Spectrum

Let us imagine a fixed attracting centre (nucleus) with the positive charge Ze. We may think of e as the elementary charge, approximately $4{\cdot}78 \times 10^{-10}$ ordinary electrostatic units, and of Z as a positive integer, the atomic number. In the field of this nucleus is an electron with the charge e'

$$(e' = -4{\cdot}78 \times 10^{-10} \text{ O.E.S. units}).$$

The potential energy of such a system is

$$V = \frac{Zee'}{r},$$

if r is the radial distance from nucleus to planetary electron. Or

$$V = -\frac{Ze^2}{r}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (42{\cdot}5)$$

Schroedinger's equation for this system is therefore

$$\nabla^2 a + \frac{8\pi^2 m}{h^2}\left(E + \frac{Ze^2}{r}\right)a = 0, \quad . \quad . \quad . \quad (42{\cdot}51)$$

m being the mass of the planetary particle. This equation can be written (cf. **42·32**)

$$\frac{\partial^2 a}{\partial r^2} + \frac{2}{r}\frac{\partial a}{\partial r} + \frac{1}{r^2}\frac{\partial^2 a}{\partial \theta^2} + \frac{\cos\theta}{r^2 \sin\theta}\frac{\partial a}{\partial \theta} + \frac{1}{r^2 \sin^2\theta}\frac{\partial^2 a}{\partial \phi^2} + \frac{8\pi^2 m}{h^2}\left(E + \frac{Ze^2}{r}\right)a = 0. \quad . \quad . \quad (42{\cdot}52)$$

We naturally seek a proper solution

$$a = R\Theta\Phi,$$

in which, as in § 42·3, Θ and Φ are respectively functions of θ and of ϕ only while R is a function of r only. On substitution in (42·52) we get

$$\Theta\Phi\frac{d^2R}{dr^2}+\frac{2}{r}\Theta\Phi\frac{dR}{dr}+\frac{1}{r^2}R\Phi\frac{d^2\Theta}{d\theta^2}+\frac{\cos\theta}{r^2\sin\theta}R\Phi\frac{d\Theta}{d\theta}$$
$$+\frac{1}{r^2\sin^2\theta}R\Theta\frac{d^2\Phi}{d\phi^2}+\left(M+\frac{B}{r}\right)R\Theta\Phi=0,$$

where

$$\left.\begin{aligned} M&=\frac{8\pi^2mE}{h^2},\\ B&=\frac{8\pi^2mZe^2}{h^2}.\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (42\cdot53)$$

Therefore

$$\frac{1}{R}\frac{d^2R}{dr^2}+\frac{2}{rR}\frac{dR}{dr}+\frac{1}{r^2\Theta}\frac{d^2\Theta}{d\theta^2}+\frac{\cos\theta}{r^2\sin\theta}\cdot\frac{1}{\Theta}\cdot\frac{d\Theta}{d\theta}$$
$$+\frac{1}{r^2\sin^2\theta}\frac{1}{\Phi}\frac{d^2\Phi}{d\phi^2}+\left(M+\frac{B}{r}\right)=0.$$

Just as in the case of (42·34) we can get a particular solution by equating

$$\frac{1}{\Phi}\frac{d^2\Phi}{d\phi^2}\text{ to }-m^2, \quad . \quad . \quad . \quad . \quad . \quad (42\cdot54)$$

where m may be any constant if we are merely concerned in obtaining a solution. But the same reasoning which, in § 42·3, required m to be integral applies here. Thus

$$\frac{1}{R}\frac{d^2R}{dr^2}+\frac{2}{rR}\frac{dR}{dr}+\frac{1}{r^2\Theta}\frac{d^2\Theta}{d\theta^2}+\frac{\cos\theta}{r^2\sin\theta}\frac{1}{\Theta}\frac{d\Theta}{d\theta}$$
$$-\frac{m^2}{r^2\sin^2\theta}+\left(M+\frac{B}{r}\right)=0,$$

or

$$\frac{r^2}{R}\frac{d^2R}{dr^2}+\frac{2r}{R}\frac{dR}{dr}+r^2M+rB+\frac{1}{\Theta}\frac{d^2\Theta}{d\theta^2}$$
$$+\frac{\cos\theta}{\sin\theta}\frac{1}{\Theta}\frac{d\Theta}{d\theta}-\frac{m^2}{\sin^2\theta}=0.$$

This equation can be solved by writing

$$\frac{r^2}{R}\frac{d^2R}{dr^2}+\frac{2r}{R}\frac{dR}{dr}+r^2M+rB=L$$

and

$$\frac{1}{\Theta}\frac{d^2\Theta}{d\theta^2}+\frac{\cos\theta}{\sin\theta}\frac{1}{\Theta}\frac{d\Theta}{d\theta}-\frac{m^2}{\sin^2\theta}=-L$$

where L may be any constant if we are not concerned with the character of the solution. But the same argument which made the constant M in **(42·351)** equal to $l(l+1)$, where l is an integer, applies here and we shall write

$$L = l(l+1).$$

Consequently

$$\frac{d^2R}{dr^2} + \frac{2}{r}\frac{dR}{dr} + \left(M + \frac{B}{r} - \frac{l(l+1)}{r^2}\right)R = 0. \quad . \quad \textbf{(42·55)}$$

As a guide, let us find the asymptotic solution for $r \to \infty$. In this case **(42·55)** simplifies to

$$\frac{d^2R}{dr^2} + MR = 0$$

and a particular solution is

$$R = e^{\pm\sqrt{-M}|r}.$$

Let us first take the case where E, and therefore also M, is negative. From the classical point of view this is the case in which the electron travels round the nucleus in an elliptical orbit. Of the two possibilities in the last equation we must take

$$R = e^{-\sqrt{-M}|r},$$

so that R may approach zero as r approaches infinity, and the suggestion naturally arises that we should try to solve **(42·55)** by

$$R = e^{-\sqrt{-M}|r}v,$$

or

$$R = e^{-\alpha r}v,$$

where v is some, so far unknown, function of r and where

$$\alpha = \sqrt{-M}|. \quad . \quad . \quad . \quad . \quad \textbf{(42·551)}$$

Writing v' for dv/dr and v'' for d^2v/dr^2 we have

$$\frac{dR}{dr} = -\alpha e^{-\alpha r}v + e^{-\alpha r}v',$$

and

$$\frac{d^2R}{dr^2} = \alpha^2 e^{-\alpha r}v - 2\alpha e^{-\alpha r}v' + e^{-\alpha r}v''.$$

Hence on substitution in **(42·55)** we obtain

$$\frac{d^2v}{dr^2} + \left(\frac{2}{r} - 2\alpha\right)\frac{dv}{dr} + \left(\frac{B - 2\alpha}{r} - \frac{l(l+1)}{r^2}\right)v = 0. \quad . \quad \textbf{(42·56)}$$

To find a solution of this equation we might, as we have done before (cf. **§ 42·2**), write

$$v = b_0 + b_1 r + b_2 r^2 + b_3 r^3 + \ldots,$$

but it is better to write v in the form

$$v = b_\nu r^\nu + b_{\nu+1} r^{\nu+1} + b_{\nu+2} r^{\nu+2} + \ldots,$$

v being a certain positive integer and b_ν, $b_{\nu+1}$, suitably chosen coefficients. By differentiation we obtain

$$v' = \nu b_\nu r^{\nu-1} + (\nu+1)b_{\nu+1}r^\nu + \ldots$$

and $$v'' = (\nu-1)\nu b_\nu r^{\nu-2} + \nu(\nu+1)b_{\nu+1}r^{\nu-1} + \ldots$$

in which v' and v'' mean respectively dv/dr and d^2v/dr^2. We substitute these expressions in **(42·56)** and equate the coefficients of the powers of r to zero. The coefficient of $r^{\nu-2}$ is

$$\{\nu(\nu-1) + 2\nu - l(l+1)\}b_\nu$$

and since, by hypothesis, b_ν is different from zero

$$\nu = l \text{ or } -(l+1).$$

We are unable to adopt $\nu = -(l+1)$ since this would make v approach infinity as r approaches zero. Therefore

$$\nu = l. \quad . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(42·57)}$$

Consequently

$$v = b_l r^l + b_{l+1}r^{l+1} + b_{l+2}r^{l+2} + \ldots \quad . \quad . \quad \textbf{(42·58)}$$

The coefficient of $r^{\nu+s-1}$, i.e. of r^{l+s-1} in the power series obtained by the substitution in **(42·56)** is

$$\{(l+s)(l+s+1) + 2(l+s+1) - l(l+1)\}b_{l+s+1} + \{(B-2\alpha) - 2\alpha(l+s)\}b_{l+s}. \quad . \quad \textbf{(42·581)}$$

This must of course be equated to zero. If now we choose α so that

$$(B-2\alpha) - 2\alpha(l+s) = 0, \quad . \quad . \quad . \quad \textbf{(42·582)}$$

(cf. **§ 42·3**). The coefficient b_{l+s+1} (and indeed all succeeding coefficients) will vanish, so that the expansion for v ends with the term

$$b_{l+s}v^{l+s}.$$

It follows, therefore, that v and consequently R will be finite and continuous for all values of r.

From the condition **(42·582)** we get

$$\alpha = \frac{B}{2(l+s+1)}$$

or in consequence of **(42·53)** and **(42·551)**

$$-E = \frac{2\pi^2 mZ^2e^4}{(l+s+1)^2h^2}. \quad . \quad . \quad . \quad \textbf{(42·59)}$$

This the familiar expression **(40·84)** for the energy values of hydrogen-like atoms. It will be noticed that the quantum integer l arose in connexion with the polar angle, θ (cf. **§ 42·3**), and it is obviously confined to positive integral values, including 0. The integer, s, on the other hand, has arisen in connexion with

the radius vector, r. Thus s is the integer n_1 of **§ 40·8** while $l + 1$ can be identified with n_2, i.e. with Bohr's k. The integer l may take the values 0, 1, 2, . . . or k may take the values 1, 2, 3, . . . Now a given value of $s + l + 1$ can be built up out of various values of s and l. For example, when $s + l + 1 = 3$ we have the following possibilities

$$\begin{aligned} s &= 0 \text{ with } l = 2, \\ s &= 1 \quad ,, \quad l = 1, \\ s &= 2 \quad ,, \quad l = 0. \end{aligned}$$

Therefore several different sets of proper functions can be associated with the same value of the integer $s + l + 1$. In the old quantum theory this case was called **degenerate** (German: *entartet*; cf. **§ 40·9**).

We have just learned that in the case where E is negative (elliptical orbits in the classical theory) the energy values are discrete and indeed they are in agreement with the values derived by using the quantum conditions $\oint p_\alpha dq_\alpha = n_\alpha h$ of the older theory. When E is positive (hyperbolic orbits of the classical theory) all real and positive values of E are possible. This theoretical result, which also emerges from the quantum conditions of the older theory, has its observational counterpart in the continuous hydrogen spectrum.

§ 42·6. Orthogonality of Proper Functions

Let us write Schroedinger's amplitude equation **(42·1)** in the following way:

$$\nabla^2 a + Ca = 0,$$

where

$$C = \frac{8\pi^2 m}{h^2}(E - V).$$

If a_m and a_n are two proper functions satisfying this equation, then

$$\nabla^2 a_m = -C_m a_m$$

and

$$\nabla^2 a_n = -C_n a_n.$$

Thus

$$\begin{aligned} & a_m \nabla^2 a_n - a_n \nabla^2 a_m \\ &= (C_m - C_n) a_m a_n \\ &= \frac{8\pi^2 m}{h^2}(E_m - E_n) a_m a_n. \end{aligned}$$

Let us substitute a_m and a_n for U and V in Green's formula **(3·12)**. We thus obtain

$$\frac{8\pi^2 m}{h^2}(E_m - E_n)\iiint a_m a_n \, dx\, dy\, dz = \iint((a_m \, \mathbf{grad}\, a_n, \mathbf{dS})) - \iint((a_n \, \mathbf{grad}\, a_m, \mathbf{dS})).$$

All such functions as a_m which are associated with discrete proper values diminish at least as rapidly as $1/r$ for large values of r, where r is the radial distance from the origin of co-ordinates. If, therefore, the volume over which the integration is extended be pushed to the limit infinity the surface integral will approach the limit zero. Therefore

$$\frac{8\pi^2 m}{h^2}(E_m - E_n)\iiint a_m a_n \, dx\, dy\, dz = 0.$$

When $m \neq n$ the difference $E_m - E_n$ is not zero, consequently

$$\iiint a_m a_n \, dx\, dy\, dz = 0.$$

Now

$$\iiint a_m{}^2 \, dx\, dy\, dz$$

is obviously not zero, and the otherwise undetermined factor in each a_m is chosen so that this integral is equal to unity (cf. **§ 41·8**). Therefore finally

$$\iiint a_m a_n \, dx\, dy\, dz = 1, \; m = n, \quad 0, \; m \neq n. \quad . \quad . \quad . \quad \mathbf{(42·6)}$$

When the constant factor in each a_m is chosen in this way these functions are said to be **normed** (or, more usually, **normalized**) and we call such functions **orthogonal functions** (cf. **§ 4·2**). The simplest examples are the cosines and sines which appear, for example, in the problem of the rotator when there is only one degree of freedom—i.e. one independent variable.

§ 42·7. Schroedinger's Interpretation of the Function ψ

Each proper solution of the wave equation, ψ_n for example, is a product of a Schroedinger amplitude a_n and an exponential, $e^{-2\pi i \nu_n t}$, and obviously the ψ_n, like the a_n, are orthogonal functions. We may therefore express the general solution (or any

solution) of the wave equation as a 'Fourier' expansion in ψ_1, ψ_2 . . . ψ_n, etc. Thus

$$\psi = \alpha_0 + \alpha_1\psi_1 + \alpha_2\psi_2 + \ldots,$$

the coefficients α_0, α_1, α_2, etc., being constants so that

$$\psi = a_0 + a_1e^{-2\pi i\nu_1 t} + a_2e^{-2\pi i\nu_2 t} + \ldots$$

in which the amplitudes a are identical (apart from a constant multiplying factor) with the amplitudes a in Schroedinger's amplitude equation **(42·1)**. It is convenient to introduce the complex quantity $\bar{\psi}$ which is conjugate to ψ, namely,

$$\bar{\psi} = \bar{a}_0 + \bar{a}_1e^{2\pi i\nu_1 t} + \bar{a}_2e^{2\pi i\nu_2 t} + \ldots$$

in which $\bar{a}_0$, $\bar{a}_1$, $\bar{a}_2$, etc., are conjugate respectively to a_0, a_1, a_2, etc. On multiplication we get for $\psi\bar{\psi}$ an expression in which the individual terms contain exponentials, $e^{2\pi i\epsilon t}$ and $e^{-2\pi i\epsilon t}$, in which the ϵ include the fundamental frequencies ν_1, ν_2, ν_3, . . . ν_m, etc., and their differences $\nu_1 - \nu_2$, $\nu_1 - \nu_3$, . . . $\nu_m - \nu_n$, . . ., etc. Schroedinger very naturally identified these frequencies with spectral frequencies and thus provided a beautiful explanation of the **combination principle** of Ritz. In order to keep as close a contact as possible with classical theory, and with Maxwell's theory in particular, he supposed the electron (in the hydrogen atom, for example) to be spread continuously over the region round the nucleus and identified the product $\psi\bar{\psi}$ with the electric density. It was probably this interpretation of $\psi\bar{\psi}$ rather than reflections like those of **§ 41·8** which led to the later view that $\psi\bar{\psi}\,dx\,dy\,dz$ should be interpreted as a measure of the probability that the electron is in the volume $dx\,dy\,dz$. The frequencies which in the older quantum theory were ascribed to **transitions** from one stationary state to another now appear as oscillation frequencies of the quantity $\psi\bar{\psi}$. The term transition is of course still appropriate, but not in Schroedinger's original theory in which radiation was not regarded as a corpuscular emission, but something very closely resembling the classical conception of it.

To find an expression for the amplitude of the vibrations, parallel to the X axis, for example, in the emitted radiation Schroedinger formed the corresponding component of the **electric moment** of the atom, namely

$$\iiint x\rho\,dx\,dy\,dz,$$

or, since he identified the electric density, ρ, with $\psi\bar{\psi}$

$$\iiint x\psi\bar{\psi}\,dx\,dy\,dz,$$

and he naturally associated

$$\iiint x\psi_m\bar{\psi}_n\,dx\,dy\,dz$$

with the frequency $\nu_m - \nu_n$. The corresponding amplitude is

$$x_{mn} = \iiint xa_m\bar{a}_n\,dx\,dy\,dz. \quad . \quad . \quad . \quad \textbf{(42·7)}$$

We can still retain this formula while rejecting Schroedinger's interpretation of the meaning of $\psi\bar{\psi}$. The present interpretation of $\psi\bar{\psi}$ as a probability of course leads to the interpretation of x_{mn} as the probability of a transition from the state m to the state n. Similarly we must regard

$$x_{nm} = \iiint xa_n\bar{a}_m\,dx\,dy\,dz,$$

or its real part, as the probability of a transition in the opposite sense.

The quantities x_{mn}, suitably arranged, constitute one of the **matrices,** namely,

$$\begin{bmatrix} x_{11}, & x_{12}, & x_{13}, & \ldots & x_{1n} & \ldots \\ x_{21}, & x_{22}, & x_{23}, & \ldots & x_{2n} & \ldots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ x_{m1}, & x_{m2}, & x_{m3}, & \ldots & x_{mn} & \ldots \\ x_{m+1,\,1}, & x_{m+1,\,2}, & x_{m+1,\,3}, & \ldots & x_{m+1,\,n} & \ldots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix} \quad \textbf{(42·71)}$$

which we shall meet in the chapter on **matrix mechanics.** It is easy to see that

$$x_{mn} = \bar{x}_{nm}.$$

We have, for example, in the integral expression for x_{nm} the product $a_n\bar{a}_m$ which may be written in the form

$$c_n e^{+i\eta_n} c_m e^{-i\eta_m},$$

in which c_n and c_m are real, or

$$c_n c_m e^{i(\eta_n - \eta_m)},$$

and when we form the corresponding product in the integral x_{mn} we get

$$c_m c_n e^{i(\eta_m - \eta_n)},$$

and the two expressions are obviously one the conjugate of the other. Matrices like **42·71,** whose components have this property are called **Hermitian matrices** after the French mathematician Hermite.

§ 42·8. Selection and Polarization Rules

It has long been known that while the frequencies observed in the spectrum of the radiation emitted by an atom (or molecule) are differences of terms characteristic of the atom (or molecule) the converse is not always the case. Bohr derived selection rules by appealing to his **correspondence principle.** For example, his quantum number k (**§ 41**) or n_2 (**§ 40·8**) could only change in any transition by $+1$ or -1. A transition in which k did not change at all, or changed by some other amount than $+1$ or -1, being forbidden. Similarly, appeal to the correspondence principle in the case of a simple harmonic emitting system restricts the possible transitions to those for which the quantum integer changes by $+1$ or -1.

We shall now show, by reference to some special cases, how this question is handled by wave-mechanics. Beginning with the simple oscillator, the Schroedinger amplitude for the frequency $\nu_n - \nu_m$ is

$$x_{mn} = \int_{-\infty}^{+\infty} x a_m \bar{a}_n \, dx = \int_{-\infty}^{+\infty} x a_m a_n \, dx, \quad . \quad . \quad \textbf{(42·8)}$$

if we use the expressions for the amplitudes given in **§ 42·2** in which complex numbers are not used, so that $\bar{a}_n = a_n$. When we substitute for a_m and a_n the expressions given by the method of **§ 42·2** we have

$$x_{mn} = \int_{-\infty}^{+\infty} e^{-x^2} v_m x v_n \, dx. \quad . \quad . \quad . \quad \textbf{(42·81)}$$

Now xv_n is a polynomial in x of degree $n+1$ and we may, since v_0, v_1, v_2, etc., are polynomials of degrees 0, 1, 2, etc., respectively, expand it in the form:

$$xv_n = \alpha_0 v_0 + \alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_{n+1} v_{n+1}.$$

Thus **(42·81)** becomes

$$x_{mn} = \int_{-\infty}^{+\infty} e^{-x^2} \{\alpha_0 v_m v_0 + \alpha_1 v_m v_1 + \ldots + \alpha_n v_m v_n + \alpha_{n+1} v_m v_{n+1}\} dx.$$

All the integrals in this sum must vanish except that one (if there is one) for which the two subscripts are equal. It is clear that when $m > n$ the integral can only differ from zero when $m = n+1$. Similarly when $n > m$ the amplitude x_{mn} can only differ from zero when $n = m+1$. The only other possibility is that for which $m = n$. In this case we have an integral

extending from $-\infty$ to $+\infty$ and containing only odd powers of x. It therefore vanishes. Thus the only non-vanishing amplitudes are those for which

$$m - n = +1 \text{ or } -1,$$

or, in the language of the older quantum theory, the only possible transitions in the case of the simple harmonic oscillator are those for which n in the expression $E = (n + 1/2)h\nu$ changes by $+1$ or -1.

Turning to the rotator, we may regard it, as we did in **§ 42·3,** as a particle of mass m whose motion is constrained to be confined to the surface of a sphere. Its rectangular co-ordinates are

$$\begin{aligned} x &= r\cos\theta, \\ y &= r\sin\theta\cos\phi, \\ z &= r\sin\theta\sin\phi, \end{aligned}$$

and r is a constant. We can simplify our investigation by combining y and z in the complex expression

$$\zeta = y + iz = r\sin\theta e^{i\phi}$$

and by introducing the corresponding complexity in the factor Φ of **(42·34)** which occurs in the amplitude a. That is, we write Φ in the form

$$\Phi = B'e^{im\phi}$$

instead of

$$\Phi = B\cos(m\phi - \eta).$$

Thus the X component of the Schroedinger amplitude, $x_{l,m;\,l',m'}$, is

$$x_{l,m;\,l',m'} = \int_0^{\pi}\int_0^{2\pi} x a_{l,m}\bar{a}_{l',m'}\,d\sigma \qquad . \quad . \quad \textbf{(42·82)}$$

where $d\sigma$ is an element, namely, $r^2\sin\theta\,d\theta\,d\phi$, of the surface of the sphere of radius r. This integral contains ϕ in the form

$$\int_0^{2\pi} e^{i(m-m')\phi}\,d\phi$$

and consequently vanishes *except when* $m = m'$.

Vibrations parallel to the X axis can only be found in the emitted radiation, therefore, as a consequence of transitions in which the integer m does not change. In such transitions the emitted radiation is polarized in the sense of having its vibrations parallel to X.

For all directions perpendicular to X we have the moment

$$\zeta_{l,m;\,l',m'} = \int_0^{\pi}\int_0^{2\pi} r \sin\theta e^{\pm i\phi} a_{l,m}\bar{a}_{l',m'}\, d\sigma, \qquad (42\cdot83)$$

so that this integral involves ϕ in the form

$$\int_0^{2\pi} e^{\pm i\phi} e^{i(m-m')\phi}\, d\phi$$

or

$$\int_0^{2\pi} e^{i(m-m'\pm 1)\phi}\, d\phi.$$

Therefore $\zeta_{l,\,m;\,l',\,m'}$ vanishes except when

$$m = m' \pm 1.$$

The radiation emitted when this moment does not vanish, i.e. when $m = m' \pm 1$ will, quite obviously, be circularly polarized as viewed in the direction of the X axis or if viewed along directions perpendicular to the X axis it will appear polarized as if its vibrations were perpendicular to the X axis.

We have next to inquire what restrictions have to be imposed on l and l' in order that the moments **(42·82)** and **(42·83)** shall not vanish. Beginning with the former, we note that m must be equal to m', so that it becomes

$$x_{l,\,m;\,l',\,m} = \int_0^{\pi}\int_0^{2\pi} x a_{l,\,m}\, a_{l',\,m}\, d\sigma.$$

The part of this involving θ is

$$\int_0^{\pi} r^2 x \Theta_{l,\,m}\Theta_{l',\,m} \sin\theta\, d\theta,$$

or

$$\int_0^{\pi} r^3 \Theta_{l,\,m}\Theta_{l',\,m} \cos\theta \sin\theta\, d\theta,$$

or finally

$$\int_{-1}^{+1} y \Theta_{l,\,m}\Theta_{l',\,m}\, dy, \qquad (42\cdot84)$$

if we drop out the constant factor r^3 and represent $\cos\theta$ by y —it is of course usual to represent $\cos\theta$ by x, but in the present investigation x has already another meaning.

Now

$$\Theta_{l,m} = (1 - y^2)^{\frac{m}{2}} v_{l,m},$$

where v is the expansion **(42·371)** and $l = m + n$, where n is the exponent of y in the last term of the expansion (cf. **42·381** and **42·382**). We have therefore for **(42·84)**

$$\int_{-1}^{+1} (1 - y^2)^m v_{l,m} (y v_{l',m}) dy$$

and, assuming $l > l'$, we can deal with it exactly as we dealt with **(42·81)** and establish that all the integrals in the expansion vanish except that for which

$$l = l' + 1$$

or, if l' is greater than l,

$$l' = l + 1.$$

The case $l = l'$ obviously, as in the case of **(42·81)**, makes the integral vanish. Precisely the same result emerges when we study the other moment **(42·83)**. The only permissible transitions are therefore those for which

$$\left.\begin{array}{l} m \to m \\ m \nearrow m + 1 \\ \quad \searrow m - 1 \\ l \nearrow l + 1 \\ \quad \searrow l - 1. \end{array}\right\} \quad . \quad . \quad . \quad . \quad . \quad \textbf{(42·84)}$$

These rules **(42·84)** apply to the Kepler case also. In particular, the integer, l (which is identical with $k - 1$, where k is the angular momentum integer in Bohr's theory), must change by $+1$ or -1 (cf. **§ 41**). The integer m does not appear explicitly in **(42·59)**; but it is in fact part of l (cf. **42·381** and **42·382**). It is the integer which appears in the theory of the Zeeman effect as Sommerfeld's magnetic quantum number **(§ 41·2)**.

CHAPTER XII

MATRIX MECHANICS

§ 42·9. Observations on the Nature of Micromechanics

WE have approached the subject of the mechanics of very small scale physical systems from the starting-point of the remarkable analogy between classical dynamics and geometrical optics—regarding light for the purposes of the analogy as an undulatory phenomenon. Micromechanics emerged when we expanded the analogy beyond the limits of *geometrical* optics so that it became one between optics, in the widest sense of optics, and dynamics. This naturally involved the representation of each particular mechanical system by a partial differential equation analogous to that which describes the propagation of light (or other) waves. Hence the name wave-mechanics.

About the time when de Broglie initiated wave-mechanics and before Schroedinger made his great contributions to the subject W. Heisenberg published a description of another, and apparently totally different, form of quantum mechanics. He was inspired mainly by Bohr's earlier work and especially by his correspondence principle, but also by a principle insisted on by the Viennese physicist, Ernst Mach, which can be traced back to Newton's *hypotheses non fingo* and which forbids the introduction into a theory of symbols representing anything that cannot be regarded as *observable*.

In order to bring out the familiar result known as the **combination principle** Heisenberg introduced into his quantum calculus a rule of multiplication formally identical with the rule for multiplication of matrices, and since its further development by Born and Jordan involved the deliberate use of matrices the term **matrix mechanics** is often applied to it. Notwithstanding its very different outward form it appears to be fundamentally identical with **wave-mechanics.** The relationship between the two may fairly aptly be compared with that between the full analytical treatment of an old-fashioned dynamical problem, and the more abbreviated treatment of it by the methods of old-fashioned vector analysis, the latter corresponding to matrix mechanics.

Independently of Born and Jordan and still earlier P. A. M. Dirac developed Heisenberg's mechanics in a way essentially identical with that of Born and Jordan, but without insisting that the non-commutative products characteristic of Heisenberg's mechanics were necessarily products of matrices. Where Born and Jordan speak of matrices Dirac speaks of **' q ' numbers** which he distinguishes from ' ordinary ' or **' c ' numbers.** All the ' q ' numbers which have any significance for physical observations are in fact matrices, so that his form of mechanics is in fact matrix mechanics.

§ 43. Heisenberg's Mechanics

We can best approach matrix mechanics and appreciate its relationship to wave-mechanics by starting out from the latter. We regard every **p** and its conjugate **q** as operators which operate on a function ψ, the wave function (cf. **§ 42**). In fact, the wave equation is set up by replacing each component of **p** by the corresponding $(h/2\pi i)\partial/\partial\mathbf{q}$. Now consider the expression $(\partial/\partial q)x\psi$ which represents the result of first multiplying ψ by x and then differentiating the product partially by q. Clearly

$$\frac{\partial}{\partial q}x\psi = \left(\frac{\partial x}{\partial q}\right)\psi + x\frac{\partial \psi}{\partial q}.$$

The brackets on the right indicate that the differentiation $\partial/\partial q$ within them is confined to x. Therefore

$$\frac{\partial}{\partial q}x\psi - x\frac{\partial}{\partial q}\psi = \left(\frac{\partial x}{\partial q}\right)\psi$$

or

$$\frac{\partial}{\partial q}x - x\frac{\partial}{\partial q} \equiv \left(\frac{\partial x}{\partial q}\right). \quad . \quad . \quad . \quad . \quad \textbf{(43)}$$

If now we replace the operation $\partial/\partial q$ by its equivalent $(2\pi i/h)p$ we get

$$\frac{2\pi i}{h}(px - xp) \equiv \left(\frac{\partial x}{\partial q}\right). \quad . \quad . \quad . \quad \textbf{(43·01)}$$

This equation would have no sense if we were to regard p and x as numerical magnitudes in the common acceptation of such a description. It is because p and x in **(43·01)** are operators in a wider sense than are ' ordinary numbers ' that $px - xp$ is not in general a null symbol. The products px and xp are in general different. They do not conform to the **commutative law** of ' ordinary ' algebraic products, or, to adopt a mode of expression which has become usual, p and x do not **commute with** one another.

It must be emphasized that p and q in **(43·01)** are **conjugate.** That is to say, when q is the positional co-ordinate q_α the p means the corresponding momentum p_α, or the operator $(h/2\pi i)\partial/\partial q_\alpha$. So we have

$$\frac{2\pi i}{h}(p_\alpha x - x p_\alpha) = \frac{\partial x}{\partial q_\alpha} \quad . \quad . \quad . \quad \textbf{(43·011)}$$

and if x happens to be q_β,

$$\frac{2\pi i}{h}(p_\alpha q_\beta - q_\beta p_\alpha) = \frac{\partial q_\beta}{\partial q_\alpha} = \begin{Bmatrix} 1, & \beta = \alpha \\ 0, & \beta \neq \alpha \end{Bmatrix}. \quad . \quad \textbf{(43·02)}$$

Similarly

$$\frac{2\pi i}{h}(p_\alpha p_\beta - p_\beta p_\alpha) = 0 \quad . \quad . \quad . \quad \textbf{(43·021)}$$

whether $\beta = \alpha$ or not.

The classical analogue of **(43·011)** appears to be

$$\frac{\partial x}{\partial q_\alpha} = \frac{\partial x}{\partial q_\alpha}\frac{\partial p_\alpha}{\partial p_\alpha} - \frac{\partial x}{\partial p_\alpha}\frac{\partial p_\alpha}{\partial q_\alpha}.$$

It should be noted that the independent variables in this equation are the **p**'s and **q**'s, so that $\partial p_\alpha/\partial p_\alpha = 1$ while $\partial p_\alpha/\partial q_\alpha = 0$. We prefer to write this equation in the following way :

$$\frac{\partial x}{\partial q_\alpha} = \frac{\partial p_\alpha}{\partial p_\alpha}\frac{\partial x}{\partial q_\alpha} - \frac{\partial p_\alpha}{\partial q_\alpha}\frac{\partial x}{\partial p_\alpha}. \quad . \quad . \quad . \quad \textbf{(43·03)}$$

The expression

$$\sum_\alpha \left(\frac{\partial u}{\partial p_\alpha}\frac{\partial v}{\partial q_\alpha} - \frac{\partial u}{\partial q_\alpha}\frac{\partial v}{\partial p_\alpha} \right)$$

is **Poisson's bracket expression** for u and v. We shall therefore call the expression **(43·011)** a Poisson bracket expression, following Dirac's usage, though it is not *strictly* analogous to the classical Poisson bracket expression which involves a summation. We shall abbreviate the statement of **(43·011)** by

$$[p_\alpha x] = \frac{\partial x}{\partial q_\alpha} \quad . \quad . \quad . \quad . \quad . \quad \textbf{(43·04)}$$

so that $[p_\alpha x]$ is an abbreviating symbol for $(2\pi i/h)(p_\alpha x - x p_\alpha)$.

Similarly by analogy with **(43·03)**

$$\frac{\partial x}{\partial p_\alpha} = \frac{\partial x}{\partial p_\alpha}\frac{\partial q_\alpha}{\partial q_\alpha} - \frac{\partial x}{\partial q_\alpha}\frac{\partial q_\alpha}{\partial p_\alpha}$$

and by comparison with **(43·04)**

$$\left.\begin{aligned} [x q_\alpha] &= \frac{\partial x}{\partial p_\alpha} \\ \text{or} \quad \frac{2\pi i}{h}(x q_\alpha - q_\alpha x) &= \frac{\partial x}{\partial p_\alpha}. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad \textbf{(43·05)}$$

If in this equation we substitute q_β for x we get

$$\frac{2\pi i}{h}(q_\beta q_\alpha - q_\alpha q_\beta) = \frac{\partial q_\beta}{\partial p_\alpha} = 0. \quad . \quad . \quad \textbf{(43·06)}$$

We are only concerned in our quantum dynamical problems with functions which do not contain the time *explicitly* but only through the p_α and q_α. When this is the case in classical dynamics

$$\frac{dx}{dt} = \sum_\alpha \left\{ \frac{\partial x}{\partial q_\alpha}\frac{dq_\alpha}{dt} + \frac{\partial x}{\partial p_\alpha}\frac{dp_\alpha}{dt} \right\}$$

and therefore when we substitute for dq_α/dt and dp_α/dt the respective expressions in the canonical equations

$$\frac{dx}{dt} = \sum_\alpha \left(\frac{\partial H}{\partial p_\alpha}\frac{\partial x}{\partial q_\alpha} - \frac{\partial H}{\partial q_\alpha}\frac{\partial x}{\partial p_\alpha} \right).$$

This is the Poisson bracket expression for H and x and we are led by it to adopt the quantum dynamical equation

$$\frac{dx}{dt} = [H,\ x] = \frac{2\pi i}{h}\{Hx - xH\}. \quad . \quad . \quad \textbf{(43·07)}$$

The H in this equation is the analogue of a certain constant quantity in classical dynamics and we must consider what **constant** means in matrix mechanics. All the symbols x, H, p, q, etc., are matrices and represent observable things or **observables,** as they are now commonly termed. The typical element of the matrix x is x_{mn}, the first subscript, m, indicating the row in the matrix, while n indicates the column. In the product

$$xy = z$$

of the two matrices x and y, the rule for multiplying is expressed by

$$z_{mn} = x_{mk}y_{kn} \text{ [summation with respect to } k\text{]},$$

as we shall see in the following section, and all the matrices which concern us contain the time in the way indicated by the equation

$$x_{mn} = a_{mn}e^{2\pi i(\nu_m - \nu_n)t}$$

in which ν_m and ν_n are constants. Hence

$$\frac{dx_{mn}}{dt} = 2\pi i(\nu_m - \nu_n)x_{mn}. \quad . \quad . \quad . \quad \textbf{(43·08)}$$

We shall regard a matrix, x, as constant when dx_{mn}/dt vanishes, whatever values the subscripts m and n may have. For a constant matrix therefore

$$2\pi i(\nu_m - \nu_n)x_{mn} = 0$$

and consequently every x_{mn} for which $m \neq n$ must vanish.

Such a matrix has therefore only **diagonal elements.** We conclude that the energy H, being a constant, must be represented by such a matrix.

From **(43·07)** we have, whether x is a constant or not,

$$2\pi i(\nu_m - \nu_n)x_{mn} = \frac{2\pi i}{h}(H_{mk}x_{kn} - x_{mk}H_{kn}),$$

in which we must sum with respect to k. Since H is a diagonal matrix the summation yields

$$2\pi i(\nu_m - \nu_n)x_{mn} = \frac{2\pi i}{h}(H_{mm} - H_{nn})x_{mn}$$

and consequently

$$\nu_m - \nu_n = (H_{mm} - H_{nn})/h. \quad . \quad . \quad \textbf{(43·09)}$$

This equation expresses at once both Bohr's postulate about spectral terms (cf. **§§ 40·5** and **40·6)** and the combination principle of Ritz. It should be noted that 1 (unity), in **(43·02),** for example, is a diagonal matrix, each of whose elements is equal to the 'ordinary' number 1. It has the characteristic property that we associate with unity, namely,

$$1 \times x = x \times 1 = x,$$

where x is any other matrix.

§ 43·1. Matrices

Consider the linear equations

$$\left.\begin{aligned} \psi_1' &= a_{11}\psi_1 + a_{12}\psi_2 + a_{13}\psi_3 + \ldots \\ \psi_2' &= a_{21}\psi_1 + a_{22}\psi_2 + a_{23}\psi_3 + \ldots \\ \psi_3' &= a_{31}\psi_1 + a_{32}\psi_2 + a_{33}\psi_3 + \ldots \\ &\ldots\ldots\ldots\ldots\ldots\ldots \end{aligned}\right\} \quad . \quad . \quad \textbf{(43·1)}$$

When we abbreviate them by writing

$$\boldsymbol{\psi}' = \mathbf{a}\boldsymbol{\psi} \quad . \quad . \quad . \quad . \quad . \quad \textbf{(43·11)}$$

a is called a **matrix** and written

$$\mathbf{a} \equiv \begin{bmatrix} a_{11}, & a_{12}, & a_{13}, \ldots \\ a_{21}, & a_{22}, & a_{23}, \ldots \\ a_{31}, & a_{32}, & a_{33}, \ldots \\ \ldots & \ldots & \ldots \end{bmatrix}$$

or in some similar way, while $\boldsymbol{\psi}$ is treated as if it were a vector in a rectangular system of co-ordinates. Similarly we may write

$$\boldsymbol{\psi}'' = \mathbf{b}\boldsymbol{\psi}' = \mathbf{ba}\boldsymbol{\psi} = \mathbf{c}\boldsymbol{\psi}$$

and since

$$\psi_i' = a_{ij}\psi_j \text{ [summation with respect to } j\text{]},$$

while

$$\psi_k'' = b_{ki}\psi_i' \text{ [summation with respect to } i\text{]},$$

therefore

$$\psi_k'' = b_{ki}a_{ij}\psi_j \text{ [summation with respect to } i \text{ and } j\text{]}.$$

But

$$\psi_k'' = c_{kj}\psi_j \text{ [summation with respect to } j\text{]}.$$

We see, therefore, when we compare the last two equations, how the rule already mentioned in the preceding section is derived.

§ 43·2. States and Observables—Proper Values

In wave-mechanics the proper (*eigen*) solutions are very prominent and are in fact solutions of the partial differential equation (the so-called wave equation) or of Schroedinger's equation possessing the property of orthogonality. Such solutions have been derived in **§§ 42·2, 42·3,** and **42·5.** The general solution, ψ, which represents, in modern terminology, the **state** of the dynamical system is expressed as a 'Fourier' expansion in terms of the proper ψ's. In matrix mechanics no analytic representation of ψ and of the proper ψ's appears explicitly. Indeed, we have no differential equations at all. Instead we adapt to our purposes the formulae of classical dynamics by representing the **observables,** e.g. energy, momenta, positional co-ordinates, etc., by matrices and by replacing differential operations by certain analogues of the classical Poisson bracket expressions as explained in **§ 43.**

The proper functions of wave-mechanics are represented in matrix mechanics as if they were the components of a vector in a rectangular co-ordinate system. Thus instead of the proper functions ψ_1, ψ_2, etc., of wave-mechanics we have the components ψ_1, ψ_2 . . . of a vector. The orthogonality of the proper functions of wave-mechanics appears in matrix mechanics as the vanishing of the scalar products

$$((\psi_1\psi_2)),\ ((\psi_2\psi_3)) \text{ or } ((\psi_m\psi_n))$$

where $m \neq n$. Indeed, we sometimes speak of **proper vectors** rather than of proper functions.

It is characteristic of a proper vector that an operator merely changes its value and not its direction, and we identify the ratio of the new value to the original value with the proper value or values of the operator, i.e. of the observable. Thus the operator $\mathbf{q}$ transforms the vector $\boldsymbol{\psi}$ into a new vector $\boldsymbol{\psi}'$ as follows:

$$\begin{aligned}
\psi_1' &= q_{11}\psi_1 + q_{12}\psi_2 + q_{13}\psi_3 + \ldots \\
\psi_2' &= q_{21}\psi_1 + q_{22}\psi_2 + q_{23}\psi_3 + \ldots \\
\psi_3' &= q_{31}\psi_1 + q_{32}\psi_2 + q_{33}\psi_3 + \ldots \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots
\end{aligned}$$

and if it happens that ψ' is equal to $c\psi$, where c is an 'ordinary' number, we obtain

$$\begin{aligned}(q_{11}-c)\psi_1 + q_{12}\psi_2 + q_{13}\psi_3 + \ldots &= 0,\\ q_{21}\psi_1 + (q_{22}-c)\psi_2 + q_{23}\psi_3 + \ldots &= 0,\\ q_{31}\psi_1 + q_{32}\psi_2 + (q_{33}-c)\psi_3 + \ldots &= 0,\\ \ldots\quad\ldots\quad\ldots\quad\ldots\quad\ldots\quad\ldots\quad\ldots\quad\ldots\quad\ldots\end{aligned}$$

Hence we find for c the equation

$$\begin{vmatrix} q_{11}-c, & q_{12}, & q_{13}, & \ldots \\ q_{21}, & q_{22}-c, & q_{23}, & \ldots \\ q_{31}, & q_{32}, & q_{33}-c, & \ldots \end{vmatrix} = 0. \qquad (43\cdot2)$$

The values of c which this equation yields are the **proper** values of the operator q. It will be noticed that when all the elements of the matrix **q**, except the diagonal ones, vanish the diagonal elements represent the proper values. We have already noticed that this is so in the case of the energy matrix **H**.

§ 43·3. Angular Momentum—Pauli's Matrices

As an illustration of matrix mechanics we begin with the rotator whose axis is fixed. Just as in **§ 41·9** we make use of the fact that the configuration of the system is repeated when the angle, θ, increases by $n2\pi$ where n is any integer.

If we replace x in **(43·011)** by $e^{in\theta}$ and q_α and p_α by θ and p respectively, we shall have

$$\frac{2\pi i}{h}(pe^{in\theta} - e^{in\theta}p) = nie^{in\theta}.$$

Each side of this equation is a matrix and we may equate the corresponding elements—for example, the element distinguished by the subscripts a, b. Thus

$$\frac{2\pi i}{h}\{p_{ac}(e^{in\theta})_{cb} - (e^{in\theta})_{ac}p_{cb}\} = ni(e^{in\theta})_{ab}.$$

The observable p (angular momentum) being a constant can be represented by a diagonal matrix. Therefore only the elements p_{aa} and p_{bb} can appear in the summation with respect to c. Consequently we get

$$\frac{2\pi}{h}(p_{aa} - p_{bb}) = n,$$

or

$$p_{aa} - p_{bb} = nh/2\pi.$$

Thus we may conclude that

$$\mathbf{p} = (n + \alpha)h/2\pi$$

where α is a constant and n is an integer. Positive and negative

values of n in this equation correspond to opposite senses of rotation and it is clear that equal numerical values of n must represent equal numerical values of p. It follows therefore that

$$(n + \alpha)h/2\pi = -(-n + \alpha)h/2\pi.$$

Hence $\alpha = 0$ and we get finally the result we have already found **(§ 41·9)** by wave-mechanical methods, namely,

$$p = nh/2\pi.$$

While we are dealing with angular momentum we might consider the case of a particle moving round a fixed centre under the influence of a force directed to or from it (cf. **§ 5·4**). The classical expression for the angular momentum **M** of the particle is

$$\mathbf{M} = \mathbf{r} \times \mathbf{p}$$

in which **p** means the ordinary translational momentum of the particle, **r** means its distance from the fixed point, and the symbol **x** is employed to indicate a vector product, since we now require the brackets [ ] to represent the Poisson bracket expressions of **§ 43.**

Such Poisson bracket expressions as $[M_x, M_y]$ are of special interest.

$$\begin{aligned}[M_x, M_y] &= [yp_z - zp_y, zp_x - xp_z],\\ &= [yp_z, zp_x] - [yp_z, xp_z] - [zp_y, zp_x] + [zp_y, xp_z],\\ &= [yp_z, zp_x] + [zp_y, xp_z]\end{aligned}$$

since the brackets with the negative sign have commuting factors. Thus, for example,

$$\begin{aligned}[yp_z, xp_z] &= \frac{2\pi i}{h}\{yp_z xp_z - xp_z yp_z\}\\ &= \frac{2\pi i}{h}\{(yx - xy)p_z{}^2\} = 0.\end{aligned}$$

Therefore

$$\begin{aligned}[M_x, M_y] &= yp_x[p_z, z] + p_y x[z, p_z],\\ &= (yp_x - xp_y)[p_z, z].\end{aligned}$$

and

$$[p_z, z] = 1$$

by **(43·02)**. So finally

$$\left.\begin{aligned}[M_x, M_y] &= -M_z\\ [M_y, M_z] &= -M_x,\\ [M_z, M_x] &= -M_y.\end{aligned}\right\} \quad \text{and similarly} \qquad \ldots\ldots \quad (43\cdot3)$$

We can also show that

$$\left.\begin{aligned}[M_x, p_y] &= -p_z,\\ [M_y, p_z] &= -p_x,\\ [M_z, p_x] &= -p_y.\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad \textbf{(43·301)}$$

Indeed,

$$[M_x, p_y] = [yp_z - zp_y, p_y] = [yp_z, p_y] - [zp_y, p_y].$$

The second of these bracket expressions vanishes since the factors of the products in it commute. Therefore

$$\begin{aligned}[M_x, p_y] &= p_z[y, p_y]\\ &= -p_z[p_y, y]\\ &= -p_z \cdot \frac{2\pi i}{h}\{p_y y - y p_y\}\\ &= -p_z.\end{aligned}$$

We might indeed have inferred the results **(43·301)** from the fact that the formal relationships of classical differential coefficients such as $\partial M_x/\partial y$ are preserved in matrix mechanics. In fact

$$-[M_x, p_y] = [p_y, M_x]$$

represents $\partial M_x/\partial y$, and since

$$\begin{aligned}M_x &= yp_z - zp_y\\ \partial M_x/\partial y &= p_z.\end{aligned}$$

We may note further that

$$\left.\begin{aligned}[M_x, y] &= -z,\\ [M_y, z] &= -x,\\ [M_z, x] &= -y,\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad \textbf{(43·302)}$$

and leave the proof to the reader.

Further

$$[\mathrm{M}_x, z^2] = [yp_z - zp_y, z^2] = [yp_z, z^2] = 2yz,$$

while

$$[M_x, y^2] = -2yz \text{ and } [M_x, x^2] = 0.$$

Consequently

$$[M_x, x^2 + y^2 + z^2] = 0$$

or

$$[M_x, \mathbf{r}^2] = 0.$$

The angular momentum components M_x, M_y, and M_z therefore commute with $\mathbf{r}^2$ and consequently with $\mathbf{r}$ and functions of $\mathbf{r}$. We can similarly show that the angular momentum components commute with $\mathbf{p}^2$ and they must therefore commute with the Hamiltonian energy expression

$$\frac{\mathbf{p}^2}{2m} + f(\mathbf{r}),$$

where $f(\mathbf{r})$ is the potential energy function of the central force.

Finally, then,

$$[H,\ M_x] = 0,$$
$$[H,\ M_y] = 0,$$
$$[H,\ M_z] = 0.$$

Otherwise expressed, **M** is a constant, as we might have otherwise anticipated.

This discussion of angular momentum as it appears in matrix mechanics will help us to describe and understand the **spin momentum** which Uhlenbeck and Goudsmit ascribed to the electron in 1925. According to them every electron has an angular momentum equal to **one half** of the unit $h/2\pi$. It belongs to the class of observables and must therefore be represented by matrices. We shall represent its components by

$$s_x h/2\pi,\ s_y h/2\pi, \text{ and } s_z h/2\pi.$$

Therefore by **(43·3)**

$$\frac{h^2}{4\pi^2}[s_x s_y] = -\, s_z \frac{h}{2\pi}$$

or

$$\frac{h^2}{4\pi^2} \cdot \frac{2\pi i}{h} \cdot \{s_x s_y - s_y s_x\} = -\, s_z \frac{h}{2\pi}$$

and therefore

$$\left.\begin{aligned} s_x s_y - s_y s_x &= i s_z \\ s_y s_z - s_z s_y &= i s_x \\ s_z s_x - s_x s_z &= i s_y. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad \textbf{(43·31)}$$

We shall for convenience use $\boldsymbol{\sigma}$ to mean $2\mathbf{s}$, so that

$$\left.\begin{aligned} \sigma_x \sigma_y - \sigma_y \sigma_x &= 2i\sigma_z \\ \sigma_y \sigma_z - \sigma_z \sigma_y &= 2i\sigma_x \\ \sigma_z \sigma_x - \sigma_x \sigma_z &= 2i\sigma_y. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad \textbf{(43·32)}$$

By hypothesis the proper values of the spin momentum components are $h/4\pi$ and $-h/4\pi$ and consequently those of each component σ_x, σ_y, or σ_z are $+1$ and -1. Consequently $\sigma_x{}^2 = \sigma_y{}^2 = \sigma_z{}^2 = 1$. If now we take one of the equations **(43·32)**, say the first one, and multiply by σ_y first on the right and then on the left and add we get

$$\sigma_y \sigma_x \sigma_y - \sigma_y \sigma_y \sigma_x + \sigma_x \sigma_y \sigma_y - \sigma_y \sigma_x \sigma_y = 2i(\sigma_y \sigma_z + \sigma_z \sigma_y)$$

or

$$\sigma_x \sigma_y{}^2 - \sigma_y{}^2 \sigma_x = 2i(\sigma_y \sigma_z + \sigma_z \sigma_y)$$

and since $\sigma_y{}^2 = 1$

$$\sigma_y \sigma_z + \sigma_z \sigma_y = 0. \quad . \quad . \quad . \quad . \quad \textbf{(43·33)}$$

Dirac describes matrices possessing this property as **anti-commuting.**

Pauli has given a set of matrices with the properties expressed by **(43·32)** and **(43·33)**. They are

$$\left.\begin{aligned} \sigma_x = \begin{bmatrix} 0, & 1 \\ 1, & 0 \end{bmatrix}, \ \sigma_y = \begin{bmatrix} 0, & -i \\ i, & 0 \end{bmatrix} \\ \text{and} \qquad \sigma_z = \begin{bmatrix} 1, & 0 \\ 0, & -1 \end{bmatrix}. \end{aligned}\right\} \quad . \quad . \quad \textbf{(43·34)}$$

These are particular cases of the more general vector in the direction (l, m, n) where l, m, and n are direction cosines:

$$\boldsymbol{\sigma} = \begin{bmatrix} n, & l - im \\ l + im, & -n \end{bmatrix}. \quad . \quad . \quad . \quad \textbf{(43·341)}$$

This matrix, it will be noticed, has the Hermitian character which the reflexions of **§ 42·7** might lead us to expect. It may easily be verified by actual multiplication that equations **(43·32)** hold for the matrices **(43·34)** and that the square of each is unity, i.e. is equal to the matrix

$$\begin{bmatrix} 1, & 0 \\ 0, & 1 \end{bmatrix}.$$

For example the 1, 1 element of the $\boldsymbol{\sigma}^2$ of **(43·341)** is

$$n^2 + (l + im)(l - im),$$

by the rule for multiplication of matrices, and this is obviously equal to 1. Similarly the 1, 2 element is

$$n(l - im) + (l - im)(-n),$$

which is equal to zero.

It will be noticed, too, that the proper values of **(43·341)** as prescribed by the determinantal equation **(43·2)** are $+1$ and -1, since the equation is

$$\begin{vmatrix} n - c, & l - im \\ l + im, & -n - c \end{vmatrix} = 0,$$

and therefore

$$-(n^2 - c^2) - (l + im)(l - im) = 0,$$

or

$$c^2 = l^2 + m^2 + n^2 = 1.$$

The fact that the square of each individual component of such a vector as $\boldsymbol{\sigma}$, namely $\sigma_x{}^2$, $\sigma_y{}^2$ and $\sigma_z{}^2$, is equal to unity, i.e. to $\boldsymbol{\sigma}^2$, or the square of the absolute value itself, perhaps needs a word or two of elucidation. The matrices of matrix mechanics, or Dirac's 'q' numbers, represent only what is observable, or what has some significance for observations. Now the usefulness of the spin hypothesis lies in the fact that it ascribes not only an angular momentum to the electron, but also lays down that its

value, i.e. its observable value, in any direction must be $h/4\pi$ multiplied by $+1$ or -1. This is perfectly expressed by such a set of matrices as **(43·34)**, the spin momentum being $\sigma h/4\pi = \mathbf{s}h/2\pi$.

One other thing deserves a little more elucidation. The M_x, M_y, etc., in **(43·3)**, for example, are apparently the sort of momenta which appear in the Hamiltonian expression for the energy of a dynamical system and consequently equation **(43·021)** would seem to apply and to suggest that $[M_x, M_y] = 0$. The explanation of this paradox is that these momenta M_x, M_y and M_z are not in fact Hamiltonian momenta.

§ 43·4. The Oscillator in Matrix Mechanics

We start with the defining equation

$$m\frac{d^2q}{dt^2} + \mu q = 0, \quad . \quad . \quad . \quad . \quad \textbf{(43·4)}$$

where m may be regarded as the mass of a particle in simple harmonic motion and μ as the restoring force per unit displacement. Now, of course, q is a matrix and

$$\frac{d^2q_{mn}}{dt^2} = -4\pi^2\nu^2{}_{mn}q_{mn}.$$

Therefore

$$(-4\pi^2\nu^2{}_{mn}m + \mu)q_{mn} = 0$$

(the subscript m must not be confused with the mass m), and since

$$\mu = 4\pi^2\nu_0{}^2m,$$

where ν_0 is the classical frequency of oscillation,

$$4\pi^2m(\nu_0{}^2 - \nu^2{}_{mn})q_{mn} = 0$$

and thus

$$\nu_{mn} = +\nu_0 \text{ or } -\nu_0. \quad . \quad . \quad . \quad . \quad \textbf{(43·41)}$$

Let us next make use of **(43·02)**, namely,

$$\frac{2\pi i}{h}(p_{mk}q_{km} - q_{mk}p_{km}) = 1 \text{ [summation with respect to } k\text{]}.$$

It should be remembered that in this equation 1 means the m, m element of the matrix 1 and that therefore the left-hand side of the equation is the ' ordinary ' number 1.

Now

$$p_{mk} = m\frac{dq_{mk}}{dt}$$

or
$$p_{mk} = 2\pi i\nu_{mk}mq_{mk}.$$

Therefore

$$-\frac{4\pi^2m}{h}\{\nu_{mk}q_{mk}q_{km} - \nu_{km}q_{mk}q_{km}\} = 1. \quad . \quad (43\cdot42)$$

There are only two values for each ν, namely ν_0 and $-\nu_0$, and in the summation with respect to k it is open to us to label the ν's so that

$$\nu_{m,\,m-1} = \nu_0$$
$$\nu_{m,\,m+1} = -\nu_0.$$

This is consistent with

$$\nu_{mk} = -\nu_{km}.$$

We get from **(43·42)** therefore

$$-\frac{4\pi^2m}{h}.2\nu_{mk}q_{mk}q_{km} = 1$$

and on summing

$$-\frac{8\pi^2m}{h}\{\nu_{m,\,m-1}\,q_{m,\,m-1}\,q_{m-1,\,m} + \nu_{m,\,m+1}\,q_{m,\,m+1}\,q_{m+1,\,m}\} = 1,$$

and finally

$$\frac{8\pi^2m\nu_0}{h}\{q_{m,\,m+1}\,q_{m+1,\,m} - q_{m,\,m-1}\,q_{m-1,\,m}\} = 1.$$

We have therefore the following sequence of equations:

$$q_{n,\,n+1}\,q_{n+1,\,n} - q_{n-1,\,n}\,q_{n,\,n-1} = \frac{h}{8\pi^2m\nu_0},$$
$$q_{n-1,\,n}\,q_{n,\,n-1} - q_{n-2,\,n-1}\,q_{n-1,\,n-2} = \frac{h}{8\pi^2m\nu_0},$$
. . . . . .
$$q_{1,\,2}\,q_{2,\,1} - q_{0,\,1}\,q_{1,\,0} = \frac{h}{8\pi^2m\nu_0},$$
$$q_{0,\,1}\,q_{1,\,0} - q_{-1,\,0}\,q_{0,\,-1} = \frac{h}{8\pi^2m\nu_0}.$$

On adding up these equations we get

$$q_{n,\,n+1}\,q_{n+1,\,n} - q_{-1,\,0}\,q_{0,\,-1} = \frac{(n+1)h}{8\pi^2m\nu_0}. \quad . \quad (43\cdot43)$$

Now we turn to the integrated equation

$$\frac{p^2}{2m} + \frac{\mu}{2}q^2 = E,$$

where E is the energy of the oscillator. We get obviously

$$\frac{p_{nk}p_{kn}}{2m} + \frac{\mu}{2}q_{nk}q_{kn} = E_{nn},$$

and therefore

$$\frac{2\pi i\nu_{nk}mq_{nk}2\pi i\nu_{kn}mq_{kn}}{2m} + \frac{\mu}{2}q_{nk}q_{kn} = E_{nn},$$

$$2\pi^2m\{\nu_0{}^2 - \nu_{nk}\nu_{kn}\}q_{nk}q_{kn} = E_{nn},$$

$$2\pi^2m\{\nu_0{}^2 + \nu_{nk}{}^2\}q_{nk}q_{kn} = E_{nn}.$$

On carrying out the summation with respect to k

$$q_{n+1,\,n}\,q_{n,\,n+1} + q_{n,\,n-1}\,q_{n-1,\,n} = \frac{E_{nn}}{4\pi^2m\nu_0{}^2}.$$

This becomes, if we make use of **(43·43)** and take $q_{-1,\,0}\,q_{0,\,-1}$ to be zero,

$$\frac{(n+1)h}{8\pi^2m\nu_0} + \frac{nh}{8\pi^2m\nu_0} = \frac{E_{nn}}{4\pi^2m\nu_0{}^2}$$

and finally

$$E_{nn} = (n + \tfrac{1}{2})h\nu_0. \quad . \quad . \quad . \quad . \quad \textbf{(43·44)}$$

It is very striking that two methods outwardly so different as the wave-mechanical one and the one just given should lead to an identical result.

CHAPTER XIII

THE QUANTUM THEORY OF THE ELECTRON

§ 43·5. The Wave Equation of the Electron

WE have seen in **§ 35** that the sum of the squares of the components, in a rectangular system of co-ordinates, of the momenta of a particle is equal to $-m_0{}^2c^2$, where m_0 is the mass of the particle when at rest. In fact, in the special theory of relativity the absolute value of the momentum of a particle is m_0ic. Thus we may write

$$p_x{}^2 + p_y{}^2 + p_z{}^2 + p_w{}^2 = -m_0{}^2c^2 \quad . \quad . \quad \textbf{(43·5)}$$

where $p_x = mv_x$ and $p_w = mic$, or if we agree to represent m_0c by p_0,

$$p_x{}^2 + p_y{}^2 + p_z{}^2 + p_w{}^2 + p_0{}^2 = 0 \quad . \quad . \quad \textbf{(43·51)}$$

(cf. **§ 36·1**).

We pass over into wave-mechanics when we regard each **p** as a differential operator. Thus

$$(p_x{}^2 + p_y{}^2 + p_z{}^2 + p_w{}^2 + p_0{}^2)\psi = 0. \quad . \quad \textbf{(43·52)}$$

The p_x, p_y, etc., in this equation symbolize the operators

$$\left.\begin{aligned} p_x &\equiv \frac{h}{2\pi i}\frac{\partial}{\partial x}, \\ p_y &= \frac{h}{2\pi i}\frac{\partial}{\partial y}, \\ p_z &= \frac{h}{2\pi i}\frac{\partial}{\partial z}, \\ p_w &= \frac{h}{2\pi i}\frac{\partial}{\partial w}, \\ p_0 &= \frac{h}{2\pi i}\frac{\partial}{\partial o}, \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad \textbf{(43·53)}$$

therefore **(43·52)** may be written

$$\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2} + \frac{\partial^2\psi}{\partial w^2} + \frac{\partial^2\psi}{\partial o^2} = 0. \quad . \quad \textbf{(43·531)}$$

Now $\partial/\partial w \equiv \partial/ic\partial t$ and

$$\frac{\partial}{\partial o} \equiv \frac{2\pi i m_0 c}{h}.$$

Therefore

$$\frac{\partial^2\psi}{\partial o^2} = -\frac{4\pi^2 m_0{}^2 c^2}{h^2}\psi,$$

and if we confine our attention to those proper solutions which contain the time in the factor $e^{-2\pi i\nu t}$—they are in fact the only ones that concern us—we have

$$\frac{\partial^2\psi}{\partial t^2} = -4\pi^2\nu^2\psi.$$

When we combine these two equations we get

$$\frac{\partial^2\psi}{\partial o^2} = \frac{m_0{}^2 c^2}{h^2\nu^2}\frac{\partial^2\psi}{\partial t^2}.$$

Consequently equation **(43·531)** becomes

$$\nabla^2\psi - \frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2} + \frac{m_0{}^2 c^2}{h^2\nu^2}\frac{\partial^2\psi}{\partial t^2} = 0.$$

If we replace $h^2\nu^2$ by the square of the energy of the particle, namely m^2c^4, we get

$$\nabla^2\psi - \frac{1}{c^2}\left(1 - \frac{m_0{}^2}{m^2}\right)\frac{\partial^2\psi}{\partial t^2} = 0$$

and since

$$\frac{m_0{}^2}{m^2} = \frac{1}{\gamma^2} = 1 - \frac{v^2}{c^2},$$

$$\nabla^2\psi - \frac{v^2}{c^4}\frac{\partial^2\psi}{\partial t^2} = 0. \quad . \quad . \quad . \quad . \quad \textbf{(43·54)}$$

This equation describes a propagation with the phase velocity c^2/v in accordance with de Broglie's prescription for the relationship between phase and group (or particle) velocity.

It is worthy of notice that equation **(43·531)** can be written in the form [1]

$$\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2} + \frac{\partial^2\psi}{\partial o^2} - \frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2} = 0 \quad . \quad \textbf{(43·55)}$$

since we may replace w by ict. This means that we may represent the wave as travelling through a 4-dimensional continuum with the same phase velocity, c, as that with which light travels through space. The suggestion therefore arises that we should

[1] Cf. Wilson and Cattermole, *Phil. Mag.*, Ser. 7, Vol. XXVII, p. 84.

regard electrons, positrons (which are electrons with their charge reversed in sign) and photons as different aspects of one and the same thing. We may think of a positron as having the momentum $+ m_0c$ along the axis o, an electron as having the momentum $- m_0c$ and a photon as having the momentum zero. The corresponding components of particle and phase velocities will be $v_0 = \pm m_0c/m = \pm c/\gamma$ and $u_0 = \pm \gamma c$ for positron or electron.

§ 43·6. Dirac's Theory of the Electron

Equation **(43·54)** and the de Broglie relationship have been convincingly verified by experiments on electron waves. Electrons, however, appear to have properties of which this equation gives no hint. There seems to be little doubt that they possess angular momentum with which is associated a magnetic moment (cf. **§ 43·3**). Dirac has attempted to account for these properties of the electron by representing the quadratic operator in **(43·52)** as a product of two factors :

$$(\alpha_x p_x + \alpha_y p_y + \alpha_z p_z + \alpha_0 p_0 + mc) \times (\alpha_x p_x + \alpha_y p_y + \alpha_z p_z + \alpha_0 p_0 - mc)\psi = 0. \quad \textbf{(43·6)}$$

The α's are matrices with the properties

$$\left.\begin{aligned} &\alpha_x^2 = \alpha_y^2 = \ldots = 1 \\ &\alpha_x\alpha_y + \alpha_y\alpha_x = 0 \\ &\alpha_y\alpha_z + \alpha_z\alpha_y = 0 \end{aligned}\right\} \quad \ldots \quad \textbf{(43·601)}$$

and so on.

Each of the two factors in this equation is equivalent to zero, so that we have two expressions for the Hamiltonian energy operator, since $H = mc^2 = c \times mc$. These are

$$\left.\begin{aligned} H &= -c(\alpha_x p_x + \alpha_y p_y + \alpha_z p_z + \alpha_0 p_0), \\ H &= +c(\alpha_x p_x + \alpha_y p_y + \alpha_z p_z + \alpha_0 p_0), \end{aligned}\right\} \quad \textbf{(43·61)}$$

so that **positive** and **negative energy states** are characteristic of Dirac's theory. It is natural to associate these energy states with electrons with a negative and with a positive charge, or in present-day terminology, with **electrons** and **positrons.** The latter had not been discovered when Dirac published his theory, and he associated the ' negative ' energy with the **proton.**

Dirac derives a magnetic moment for his electron in the following way. He replaces the p_x, p_y and p_z of **(43·6)** by $p_x + eK_x$, $p_y + eK_y$ and $p_z + eK_z$ respectively and mc^2 by $mc^2 + eV$ or mc by $mc + \dfrac{eV}{c}$ where V is the scalar potential and

K_x, K_y and K_z are the components of the vector potential as defined by **(36·1)**. The factors of **(43·6)** now become

$$\{\alpha_x(p_x + eK_x) + \alpha_y(p_y + eK_y) + \alpha_z(p_z + eK_z) + \alpha_0 m_0 c + \left(mc + \frac{eV}{c}\right)\} \times \left\{\alpha_x(p_x + eK_x) + \alpha_y(p_y + eK_y) + \alpha_z(p_z + eK_z) + \alpha_0 m_0 c - \left(mc + \frac{eV}{c}\right)\right\}\psi = 0. \quad \textbf{(43·62)}$$

The multiplication of the two factors yields, in addition to terms like $\alpha_x{}^2(p_x + eK_x)^2$, $\alpha_y{}^2(p_y + eK_y)^2$, etc., products which go together in pairs, as for example

$$\alpha_x\alpha_y(p_x + eK_x)(p_y + eK_y)\psi + \alpha_y\alpha_z(p_y + eK_y)(p_x + eK_x)\psi.$$

The sum of these two products, since $\alpha_y\alpha_x = -\alpha_x\alpha_y$, is

$$\alpha_x\alpha_y\{p_xp_y + ep_xK_y + eK_xp_y + e^2K_xK_y - p_yp_x - ep_yK_x - eK_yp_x - e^2K_yK_x\}\psi. \quad \textbf{(43·63)}$$

Now p_x and p_y are the operators $(h/2\pi i)\partial/\partial x$ and $(h/2\pi i)\partial/\partial y$ and they commute in the products p_xp_y and p_yp_x. The components K_x and K_y also commute in the products $e^2K_xK_y$ and $e^2K_yK_x$. The sum **(43·63)** therefore reduces to

$$\alpha_x\alpha_y\left\{\frac{eh}{2\pi i}\frac{\partial}{\partial x}K_y\psi + eK_x\frac{h}{2\pi i}\frac{\partial}{\partial y}\psi - \frac{eh}{2\pi i}\frac{\partial}{\partial y}K_x\psi - eK_y\frac{h}{2\pi i}\frac{\partial}{\partial x}\psi\right\}$$

and this is equivalent to

$$\alpha_x\alpha_y\frac{eh}{2\pi i}\left(\frac{\partial K_y}{\partial x} - \frac{\partial K_x}{\partial y}\right),$$

if we drop the symbol ψ.[1] For electrostatic units and empty space this is

$$\alpha_x\alpha_y\frac{eh}{2\pi ic}H_z,$$

to which we must add

$$\alpha_y\alpha_z\frac{eh}{2\pi ic}H_x$$

and

$$\alpha_z\alpha_x\frac{eh}{2\pi ic}H_y.$$

We have further the sums of products like

$$\alpha_x(p_x + eK_x)\left(mc + \frac{eV}{c}\right)\psi - \alpha_x\left(mc + \frac{eV}{c}\right)(p_x + eK_x)\psi$$

[1] In the foregoing expression the operator's $\partial/\partial x$ and $\partial/\partial y$ do not operate on ψ, the differentiations $\partial\psi/\partial x$ and $\partial\psi/\partial y$ having already been carried out.

which can be seen to reduce to

$$-\alpha_x \frac{eh}{2\pi ic}\left\{-\frac{\partial V}{\partial x}-\frac{\partial K_x}{\partial t}\right\},$$

when we remember that

$$mic \equiv \frac{h}{2\pi i}\frac{\partial}{\partial w},$$

or

$$mc \equiv -\frac{h}{2\pi ic}\frac{\partial}{\partial t};$$

or finally

$$-\alpha_x \frac{eh}{2\pi ic}\mathcal{E}_x,$$

to which we must add

$$-\alpha_y \frac{eh}{2\pi ic}\mathcal{E}_y$$

and

$$-\alpha_z \frac{eh}{2\pi ic}\mathcal{E}_z.$$

Thus the product **(43·62)** becomes

$$\{(p_x+eK_x)^2+(p_y+eK_y)^2+(p_z+eK_z)^2-\left(mc+\frac{eV}{c}\right)^2$$
$$+\frac{eh}{2\pi ic}(\alpha_x\alpha_y H_z+\alpha_y\alpha_z H_x+\alpha_z\alpha_x H_y)$$
$$-\frac{eh}{2\pi ic}(\alpha_x\mathcal{E}_x+\alpha_y\mathcal{E}_y+\alpha_z\mathcal{E}_z)$$
$$+m_0{}^2c^2\}\psi=0. \quad . \quad . \quad . \quad . \quad \textbf{(43·64)}$$

The proper values of the product matrices $\alpha_x\alpha_y$, etc., are $\sqrt{-1}$ and $-\sqrt{-1}$, while those of α_x, α_y, α_z, etc., are $+1$ and -1, as we shall see later; so that the terms containing H_x, H_y and H_z are real, while those containing $\mathcal{E}_x$, $\mathcal{E}_y$ and $\mathcal{E}_z$ are imaginary. Dirac claims that no physical meaning can be ascribed to these latter.

Now

$$mc+\frac{eV}{c}=\frac{E}{c}+\frac{eV}{c},$$

where E is the energy of the particle itself. Further,

$$E=m_0c^2+E_1$$

where E_1 is the kinetic energy of the particle and we may assume (if we suppose the kinetic energy of the electron to be relatively small) that

$$m_0c^2 >> E_1.$$

Thus

$$\left(mc + \frac{eV}{c}\right)^2 = (m_0c^2 + E_1 + eV)^2/c^2$$
$$= m_0{}^2c^2 + 2m_0E_1 + 2m_0eV$$

and consequently we get from **(43·64)**

$$E_1 = -eV + \frac{1}{2m_0}\{(p_x + eK_x)^2 + (p_y + eK_y)^2 + (p_z + eK_z)^2\}$$
$$+ \frac{eh}{4\pi m_0 c}\{(-i\alpha_x\alpha_y)H_z + (-i\alpha_y\alpha_z)H_x + (-i\alpha_z\alpha_x)H_y\}$$
$$+ \frac{eh.i}{4\pi m_0 c}\{\alpha_x\mathcal{E}_x + \alpha_y\mathcal{E}_y + \alpha_z\mathcal{E}_z\}. \quad . \quad . \quad . \quad . \quad \textbf{(43·65)}$$

If we define three matrices σ_x, σ_y and σ_z to have the meanings

$$\left.\begin{aligned}\sigma_x &= -i\alpha_y\alpha_z,\\ \sigma_y &= -i\alpha_z\alpha_x,\\ \sigma_z &= -i\alpha_x\alpha_y,\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad \textbf{(43·66)}$$

we may write the terms containing H in the form

$$\frac{eh}{4\pi m_0 c}\{\sigma_xH_x + \sigma_yH_y + \sigma_zH_z\}$$

or

$$\frac{eh}{4\pi m_0 c}((\boldsymbol{\sigma}\mathbf{H})). \quad . \quad . \quad . \quad . \quad . \quad \textbf{(43·67)}$$

This term represents part of the energy of the electron, so that we must ascribe to it a magnetic moment

$$\boldsymbol{\sigma}\frac{eh}{4\pi m_0 c}. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(43·671)}$$

We shall see in the next section that $\boldsymbol{\sigma}$ has the real proper values $+1$ and -1 and we conclude that the electron has a magnetic moment of plus or minus a Bohr magneton. This means that an observation of the value of the component of the electron's magnetic moment in any direction, if it could be made, would yield either $+\frac{eh}{4\pi m_0 c}$ or $-\frac{eh}{4\pi m_0 c}$.

§ 43·7. The Matrices $\boldsymbol{\alpha}$

In describing the matrices $\boldsymbol{\alpha}$ it is convenient to use the picture of a particle moving through a 4-dimensional region made up of space plus the extra dimension O (cf. **43·55**), or, better still, to picture it in a 5-dimensional continuum like that of Kaluza, just as we picture a photon in the Galilean space-time of Einstein

(cf. **43·531**). In doing this we are not necessarily making use of Kaluza's relativity theory, but gaining the advantage of the *Anschaulichkeit* of a geometrical representation. Let us think of the continuum as possessing a Euclidean character and l, m, n, j and k as the direction cosines of the direction of some vector in the continuum referred to rectangular axes of co-ordinates X, Y, Z, W and O respectively. We may associate the following matrix **(43·7)**, and also others, with this direction:

$$\begin{bmatrix} k, & 0, & n-ij, & l-im \\ 0, & k, & l+im, & -n-ij \\ n+ij, & l-im, & -k, & 0 \\ l+im, & -n+ij, & 0 & -k \end{bmatrix} \quad . \quad (43\cdot7)$$

If we take in succession the directions of the co-ordinate axes, $l=1$, $m=1$, $n=1$, $j=1$ and $k=1$ respectively, we get matrices which have the properties assigned to α_x, α_y, α_z, α_w and α_o. Of these α_o is a diagonal matrix while α_w is an additional one not used in the foregoing presentation of Dirac's factors, but which might have been used with the mc **(43·6)** instead of $+1$ and -1. It is obtained, as already indicated, by writing $j=1$, $l=m=n=k=0$, in **(43·7)**. Thus

$$\alpha_w = \begin{bmatrix} 0, & 0, & -i, & 0 \\ 0, & 0, & 0, & -i \\ i, & 0, & 0, & 0 \\ 0, & i, & 0, & 0 \end{bmatrix} \quad . \quad . \quad . \quad (43\cdot71)$$

It is easy to see by using the rule for multiplying matrices that the products $\alpha_x\alpha_y$, $a_y\alpha_z$, etc., have the proper values $+i$ and $-i$. In fact

$$\left.\begin{aligned} \alpha_x\alpha_y &= \begin{bmatrix} i, & 0, & 0, & 0 \\ 0, & -i, & 0, & 0 \\ 0, & 0, & i, & 0 \\ 0, & 0, & 0, & -i \end{bmatrix} \\ \alpha_y\alpha_z &= \begin{bmatrix} 0, & i, & 0, & 0 \\ i, & 0, & 0, & 0 \\ 0, & 0, & 0, & i \\ 0, & 0, & i, & 0 \end{bmatrix} \end{aligned}\right\} \quad . \quad . \quad (43\cdot72)$$

and

$$\left.\alpha_z\alpha_x = \begin{bmatrix} 0, & 1, & 0, & 0 \\ -1, & 0, & 0, & 0 \\ 0, & 0, & 0, & 1 \\ 0, & 0, & -1, & 0 \end{bmatrix}\right\}$$

The matrices σ_x, σ_y and σ_z **(43·66)** are therefore expressed by

$$\sigma_x = \begin{bmatrix} 0, & 1, & 0, & 0 \\ 1, & 0, & 0, & 0 \\ 0, & 0, & 0, & 1 \\ 0, & 0, & 1, & 0 \end{bmatrix}, \quad \sigma_y = \begin{bmatrix} 0, & -i, & 0, & 0 \\ i, & 0, & 0, & 0 \\ 0, & 0, & 0, & -i \\ 0, & 0, & i, & 0 \end{bmatrix},$$

$$\sigma_z = \begin{bmatrix} 1, & 0, & 0, & 0 \\ 0, & -1, & 0, & 0 \\ 0, & 0, & 1, & 0 \\ 0, & 0, & 0, & -1 \end{bmatrix} \quad . \quad . \quad . \quad \textbf{(43·73)}$$

These matrices represent in fact a slight generalization of the **spin matrices** of **(43·34)**. They only differ by a sort of doubling which does not affect their proper values or their multiplicative properties.

§ 43·8. The Spin Momentum

The angular momentum expressed by the vector product $\mathbf{r} \times \mathbf{p}$ (which we formerly wrote $[\mathbf{r}, \mathbf{p}]$) is constant under the central forces of the classical theory and this constancy necessarily passes over into matrix mechanics, provided we use a Hamiltonian operator of the classical form, i.e. a function of the radial distance from the centre of force and of the *squares* of the momentum components. When, however, we adopt the Hamiltonian operators of **(43·61)**, as we must if we accept Dirac's theory, which are *linear* in the momentum components, the angular momentum $\mathbf{r} \times \mathbf{p}$ is no longer constant. Instead of this we find that

$$\boxed{\mathbf{M} + \boldsymbol{\sigma} h/4\pi = \text{constant.}} \quad . \quad . \quad . \quad \textbf{(43·8)}$$

This can be shown in the following way: First of all

$$\frac{dM_x}{dt} = [H, M_x]$$

or
$$\frac{dM_x}{dt} = [c(\alpha_x p_x + \alpha_y p_y + \alpha_z p_z + \alpha_0 m_0 c + f(r), M_x]$$

where $f(r)$ is the potential energy function of the central force. So that dM_x/dt is represented by a sum of Poisson bracket expressions. Only two of them do not vanish, i.e. there are only two whose factors do not commute. Thus

$$\frac{dM_x}{dt} = c[\alpha_y p_y, M_x] + c[\alpha_z p_z, M_x].$$

Therefore, by **(43·301)**,

$$\frac{dM_x}{dt} = c(\alpha_y p_z - \alpha_z p_y). \quad . \quad . \quad . \quad . \quad \textbf{(43·81)}$$

Let us next consider the rate of increase of $\alpha_y\alpha_z$. We are using the phrase 'rate of increase' in the Pickwickian sense of matrix mechanics. We have

$$\frac{d(\alpha_y\alpha_z)}{dt} = [H,\ \alpha_y\alpha_z],$$

H being the Hamiltonian operator used to obtain **(43·81)**. This reduces, when we exclude the parts of H which commute with $\alpha_y\alpha_z$, to

$$\frac{d(\alpha_y\alpha_z)}{dt} = c[\alpha_y p_y,\ \alpha_y\alpha_z] + c[\alpha_z p_z,\ \alpha_y\alpha_z].$$

Therefore

$$\frac{d(\alpha_y\alpha_z)}{dt} = \frac{2\pi ic}{h}(\alpha_y p_y \alpha_y\alpha_z - \alpha_y\alpha_z\alpha_y p_y) + \frac{2\pi ic}{h}(\alpha_z p_z\alpha_y\alpha_z - \alpha_y\alpha_z\alpha_z p_z).$$

Now Dirac supposes the α's to commute with the p's and it will be remembered that $\alpha_z\alpha_y = -\alpha_y\alpha_z$, etc. Therefore

$$\frac{d(\alpha_y\alpha_z)}{dt} = \frac{2\pi ic}{h}(\alpha_y{}^2 p_y\alpha_z + \alpha_y{}^2\alpha_z p_y) - \frac{2\pi ic}{h}(\alpha_z{}^2 p_z\alpha_y + \alpha_z{}^2\alpha_y p_z).$$

Hence

$$\frac{d\sigma_x}{dt} = \frac{d(-i\alpha_y\alpha_z)}{dt},$$

$$= \frac{4\pi c}{h}(\alpha_z p_y - \alpha_y p_z),$$

and therefore

$$\frac{d}{dt}(\sigma_x h/4\pi) = c(\alpha_z p_y - \alpha_y p_z). \quad . \quad . \quad . \quad \textbf{(43·82)}$$

On adding together **(43·81)** and **(43·82)** we get finally

$$\frac{d}{dt}\left(M_x + \frac{\sigma_x h}{4\pi}\right) = 0,$$

in agreement with **(43·8)**.

It is not therefore the angular momentum **M** which is conserved, but the sum

$$\mathbf{M} + \boldsymbol{\sigma} h/4\pi,$$

and since **M** represents an orbital angular momentum we are forced to associate $\boldsymbol{\sigma} h/4\pi$ with an angular momentum of the electron itself. This is the spin momentum.

§ 43·9. Observations on Dirac's Theory

We have tacitly assumed that the p_x, p_y, etc., of **(43·6)**, for example, have the same meanings as the respective p_x, p_y, etc., in **(43·62)**. In fact, we have assumed that in all these equations p_x, p_y, etc., are the representatives of mv_x, mv_y, etc., respectively. Now it is an essential feature of the theory that in an electromagnetic field the rôles of the momenta are taken over by the extended momenta, $\Pi_x(=p_x+eK_x)$, Π_y, etc., and it is these, and not the p_x, p_y, etc., which are conjugate to the respective positional co-ordinates x, y, . . . and which should be represented by the operators $(h/2\pi i)\partial/\partial x$, $(h/2\pi i)\partial/\partial y$, . . . respectively. Dirac's own description of his theory is not very clear on this point.

The expression

$$(p_x^2+p_y^2+p_z^2+p_w^2+p_0^2)\psi,$$

which vanishes according to the special theory of relativity, is in fact equivalent to

$$\{(\Pi_x-eK_x)^2+(\Pi_y-eK_y)^2+(\Pi_z-eK_z)^2 + m_0^2c^2-\left(\frac{E}{c}-\frac{eV}{c}\right)^2\}\psi,$$

so that with this interpretation of Dirac's theory the factors of **(43·62)** only need to be modified by changing the sign of e. The ambiguity may be described as resting on the uncertainty as to whether e means $-4{\cdot}78\times10^{-10}$ or $+4{\cdot}78\times10^{-10}$ O.E.S. units for the electron, and whatever choice we make the result will represent either an electron with negative charge or one with positive charge (positron). There is of course an associated ambiguity in the meaning of E_1 of **(43·65)** which is either the kinetic energy of the particle or the Hamiltonian energy function.

Turning back to **43·61** the energy, H, of the particle, which of course is equal to mc^2, can be **positive** or **negative.** Dirac associates the former with the electron and the latter with the positron.

In order to meet the difficulty due to the fact that the measured value of mc^2 is necessarily positive even for a positron, he makes an appeal to Pauli's **exclusion principle,** according to which only one electron (or positron) can be in a given state and supposes that nearly all the negative energy states are occupied, each one by an electron. An unoccupied negative energy state will appear as something with positive energy, and this he identifies with a positron.[1] The set of occupied states are the electrons,

[1] Originally he identified it with a proton, positrons not having been discovered at that time.

the comparatively few unoccupied states are the positrons. This picture gives a plausible reason for the comparative permanence of an electron and the short term of existence of a positron.

Cattermole and Wilson have recently outlined a theory which rests on Kaluza's relativity theory, in so far at any rate as it makes use of a 5-dimensional continuum and identifies the charge on the elementary particle—apart from a constant of suitable dimensions—with the momentum m_0c along the axis, O, (cf. **§ 43·5**). Instead of factorizing the quadratic operator **(43·52),** as does Dirac, they represent the electron (positron and photon) by differential equations which may be described as Maxwell's equations (of the electromagnetic field) generalized to take a 5-dimensional form in the spirit of Kaluza's relativity theory. We cannot describe this here; but must refer the reader to the bibliographical list at the end of this volume. This electron theory has the advantage that it starts out from premisses which are less *ad hoc* and arbitrary than are those of Dirac's theory and it yields no electric moment, not even an imaginary one.

H. T. Flint arrives at Dirac's results by starting out with a new kind of geometry in which lengths and vectors are represented from the outset by matrices. His work is based on Kaluza's 5-dimensional relativity and he assumes the path of an electron, like that of a photon, to be a null geodesic in this continuum. An interesting feature of his work is the use he makes of Weyl's ' gauge ' hypothesis. In fact, Dirac's equation may be described as a ' gauging ' equation.

CHAPTER XIV

MISCELLANEOUS SUBJECTS

§ 44. The Magnetic Quantum Number

THE mutual energy of an electron orbit and a magnetic field, according to the classical theory, is (cf. **41·25**)

$$J_\phi eH/4\pi m_0 c \quad . \quad . \quad . \quad . \quad . \quad \textbf{(44)}$$

in which J_ϕ is the azimuthal phase integral on which the old quantum conditions **(§ 40·8)** impose the condition

$$J_\phi = n_\phi h.$$

Thus the mutual energy of a magnetic field and an electron orbit is, according to the older quantum theory,

$$n_\phi(eh/4\pi m_0 c)\mathrm{H} \quad . \quad . \quad . \quad . \quad . \quad \textbf{(44·01)}$$

and the orbit has therefore a magnetic moment

$$n_\phi eh/4\pi m_0 c.$$

It is thus an integral multiple of $eh/4\pi m_0 c$ which is known as the **Bohr magneton.** Sommerfeld calls n_ϕ the **magnetic quantum number.**

§ 44·1. The Inner Quantum Number—Landé's Factor

Another quantum number, introduced by Sommerfeld, is the **inner quantum number.** Its purpose is to distinguish the members in the multiple structure of S, P, D, etc., terms. Thus, for example, in the alkali metals each P term is double **(§ 40·5).** It consists in fact of two terms differing little in value. The inner quantum numbers which distinguish them are $\frac{1}{2}$ and $\frac{3}{2}$, as will be explained later. Thus we may symbolize them by $P_{\frac{1}{2}}$ and $P_{\frac{3}{2}}$. The quantum number j, which in the case just mentioned has the values 1/2 and 3/2, was at first believed to represent the total angular momentum of the atom. It is now known to be a measure of the total angular momentum of the electron orbit, or orbits, in question, i.e. orbital + spin momentum, as illustrated in Fig. 44·1. The associated magnetic moment is not, however, in the direction of j on account of the fact that the

ratio of the electron's magnetic moment to its spin momentum is double that of the orbital magnetic moment to the associated orbital angular momentum. To obtain the component of the total magnetic moment in the direction of j we must count the contribution from the electron spin twice over. That is to say, the component of the magnetic moment in the direction of j is not $jeh/4\pi m_0 c$ but $(j + s\cos\varepsilon)(eh/4\pi m_0 c)$. It will be seen from Fig. 44·1 that

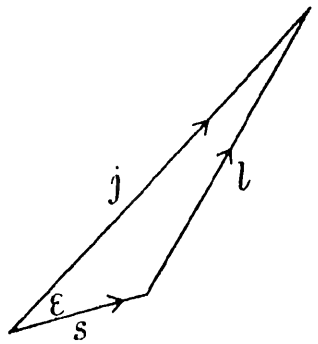

FIG. 44·1

$$j + s\cos\varepsilon = j\left(1 + \frac{j^2 + s^2 - l^2}{2j^2}\right)$$

and the wave mechanical investigations of §§ **42·3** and **42·5** naturally suggest the replacement of j^2, s^2 and l^2 by $j(j+1)$, $s(s+1)$ and $l(l+1)$ respectively. We shall therefore write

$$j + s\cos\varepsilon = j\left(1 + \frac{j(j+1) + s(s+1) - l(l+1)}{2j(j+1)}\right) \quad . \quad \textbf{(44·1)}$$

or briefly

$$j + s\cos\varepsilon = jg.$$

The factor g is known as **Landé's splitting factor** (German : *Aufspaltungsfaktor*). In this connexion it should be noted that the integer l is identified with $k - 1$, where k is Bohr's angular momentum quantum number (cf. **§ 42·5**).

The proper values of the component of j in any direction—and we are more especially interested in the direction of an applied magnetic field, **H**—must be equal to j itself or differ from it by an integer and therefore we have for the component of the magnetic moment of an electron—orbital + spin—

$$j.g.B,\ (j-1).g.B,\ \ldots\ -(j-1).g.B,\ -j.g.B \quad . \quad \textbf{(44·11)}$$

in which B represents the Bohr magneton, $eh/4\pi m_0 c$.

§ 44·2. PAULI'S PRINCIPLE AND ATOMIC STRUCTURE

The letters j, l and s are also used to distinguish or specify spectral terms and the details of their multiple structure ; but in this section they have reference to individual orbits only. The character of an orbit is completely specified when the values of n (the **total quantum number**), of l (or k) (the angular momentum quantum number), of s (which are confined to $+1/2$ or $-1/2$) and of the component of l in the direction of s. There are thus four numbers required to specify an electron orbit.

The electronic structure of atoms and, associated with it,

the peculiarities of their spectra are based in an important way on an empirical principle discovered by Pauli according to which no two electrons in an atom can be associated with identical values of these four numerical quantities.

As we have already learned, Bohr adopted Rutherford's view that a neutral atom consists of a massive (but small) nucleus with planetary electrons moving round it. The nucleus carries a positive charge which is an integral multiple of the elementary charge, the integer—now called the **atomic number**—being identical with the number of electrons. Let us first consider those electron orbits of an atom for which $n = 1$. Now $n = n_1 + k = n_1 + l + 1$ and l cannot be less than zero, so that in this case n_1 and l must be zero. Hence the component of l in the direction of s must also be zero. There can be, therefore, only *two* orbits for which $n = 1$, those for which $s = +1/2$ and $-1/2$. In the hydrogen atom there is one such orbit and two of them in the helium atom. When we pass to orbits for which $n = 2$, one of the possibilities for l is $l = 0$ and clearly we shall have two such orbits. But when $n = 2$ another possibility for l is $l = 1$. In this case the component of l in the s direction may have the values $+1$, 0 and -1 and each of these possibilities may be associated with either of the values $+1/2$ or $-1/2$ for s. There are therefore 6 orbits of this kind. Altogether, then, there are 8 orbits for which $n = 2$, and by Pauli's principle no atom may have more than 8 electron orbits for which $n = 2$.

When we proceed to orbits for which $n = 3$ the foregoing cases for which $l = 0$ and $l = 1$ are repeated and we have 8 such orbits; but there is now the possibility $l = 2$ and consequently its components in the s direction are $+2$, $+1$, 0, -1, -2. Each of these 5 possibilities can be associated with $s = +1/2$ or $s = -1/2$. There are therefore 10 of these orbits and in all 18 orbits for which $n = 3$. Bohr uses the symbol n_k to mark an orbit which has the total quantum number n and the angular momentum quantum number k $(= l + 1)$. Thus there are in an atom at most

2	orbits	of	the	1_1	type,	
2	,,	,,	,,	2_1	,,	,
6	,,	,,	,,	2_2	,,	,
2	,,	,,	,,	3_1	,,	,
6	,,	,,	,,	3_2	,,	,
10	,,	,,	,,	3_3	,,	.

We can proceed in this way indefinitely and include all the

elements that are actually known. Helium has the complete set, namely 2, of orbits for which $n = 1$. The next element is lithium which has, in addition to the two 1_1 orbits of helium, a single outer orbit which, in the unexcited state of the atom, is a 2_1 orbit. It is an orbit in the S state (since $k = 1$). Proceeding further we meet in succession beryllium, boron, carbon, nitrogen, oxygen, fluorine, which has for example two 1_1 orbits, two 2_1 orbits and five 2_2 orbits, neon which has the complete sets of orbits for which $n = 1$ and $n = 2$. After neon comes sodium with an outer electron like that of lithium, only this is in a 3_1 orbit at any rate in its unexcited state. The outermost electron system (in the case of fluorine, for example, the seven orbits for which $n = 2$ or in the case of sodium the single orbit for which $n = 3$) determines the chemical properties of the element and also the character of its optical spectra (not its X-ray spectra which are associated with the inner completed systems of electron orbits). All elements, such as helium and neon, which have only completed systems of orbits, are chemically very inert. This indicates that completed sets of orbits have great stability and we can understand the great chemical activity of an element like fluorine which needs a single extra electron orbit for completion and its affinity for an element like sodium which has such a single electron orbit in its uncompleted electron orbit system. It should be noted that the resultant of the magnetic moments of completed electron systems is zero and that the para- and ferromagnetic properties of elements are due to uncompleted systems of electron orbits.

The general plan, outlined above, of the electronic structure of atoms is almost that originally suggested by Bohr before the discovery of Pauli's principle.

§ 44·3. The Multiple Structure of Spectral Terms

The letters j, l and s are sometimes used in a wider sense than that of specifying electron orbits. The letter l distinguishes the different classes of terms. For example, S terms are associated with $l = 0$, P terms with $l = 1$, and so on. Different values of j distinguish the members in the multiple structure of the term. In the case of an element like sodium, the optical spectrum of which is associated with a single electron, the j, l and s, which are used in describing its spectral terms and their multiple structure, are identical with the j, l and s which describe the orbit. In the unexcited state of the atom $n = 3$ and $l = 0$. Its energy is that of an S term and $j = l + s = 0 + 1/2$ or $0 - 1/2$. So far as the energy is concerned we cannot distinguish between

$j = 1/2$ and $j = -1/2$. It is therefore convenient to represent this term by $3S_{1/2}$, the number 3 indicating the value of Bohr's total quantum number n.

It may be remarked in passing that all S terms are single. The corresponding terms of lithium and potassium would naturally be described by $2S_{1/2}$ and $4S_{1/2}$ respectively. When the sodium atom is excited the total quantum number of the outer electron orbit might become 4, for example, and in such a case l would be 0, 1, 2 or 3. When $l = 1$ the energy value is that of a P term. There are clearly two members of the P term, since $j = l + s$ and may therefore be $1 + \frac{1}{2}$ or $1 - \frac{1}{2}$. That is to say, $j = 3/2$ or $1/2$. The D term is associated with $l = 2$ and its two members are associated with $j = 2\frac{1}{2}$ and $j = 1\frac{1}{2}$.

When we turn to atoms like that of magnesium, for example, which have two outer electrons, the l which distinguishes the term is the vector sum of the l's of the two electron orbits. The same remark applies to s in the equation $j = l + s$ which determines the biggest value of j. The term multiplicity is $2s + 1$. In the case of the alkali metals, as we have seen, $s = 1/2$ and the multiplicity is 2. In the case of magnesium, calcium, etc., which have two outer electron orbits $s = 1/2 + 1/2 = 1$ or $s = 1/2 - 1/2 = 0$. When $s = 1$ the number j can have three values, since it varies by integers between $l + 1$ and $l - 1$. It can in fact have the values $l + 1$, l, $l - 1$. In this case the terms are triple. But with such elements s may be zero. In this case the terms are single. All elements therefore with two outer electron orbits have triplet and also singlet systems of terms in accordance with the equation

$$\text{multiplicity} = 2s + 1$$

and the possibilities

$$s = 1/2 + 1/2 = 1,$$
$$s = 1/2 - 1/2 = 0.$$

§ 44·4. X-ray Spectra

In the first decade of this century Barkla discovered what he termed **characteristic** or **fluorescent X-radiation.** Some elements (silver, for example) he found to emit more than one sort of characteristic radiation when exposed to a sufficiently penetrating primary radiation. He distinguished these different types of characteristic radiation by their absorbability in aluminium and labelled them K radiation (the most penetrating), L radiation, M radiation, and so on. Subsequent investigations with X-ray spectrographs have revealed each of these to consist of a series of lines like an optical spectral series. The origin of

the K emission spectrum is now believed to be due to the ejection from the atom of one of the electrons (called K electrons) for which $n = 1$. The line K_α is due to an L electron ($n = 2$) taking the vacant place. Similarly the K_β line is due to an M electron ($n = 3$) taking the vacant place, and so on. The L radiation is similarly due to the ejection from the atom of an L electron ($n = 2$) and the consequent filling up of the vacancy by an M, N, O, etc., electron. The spectral terms whose differences give the frequencies of the radiation corresponding to the observed spectral line exhibit a multiplicity which can be

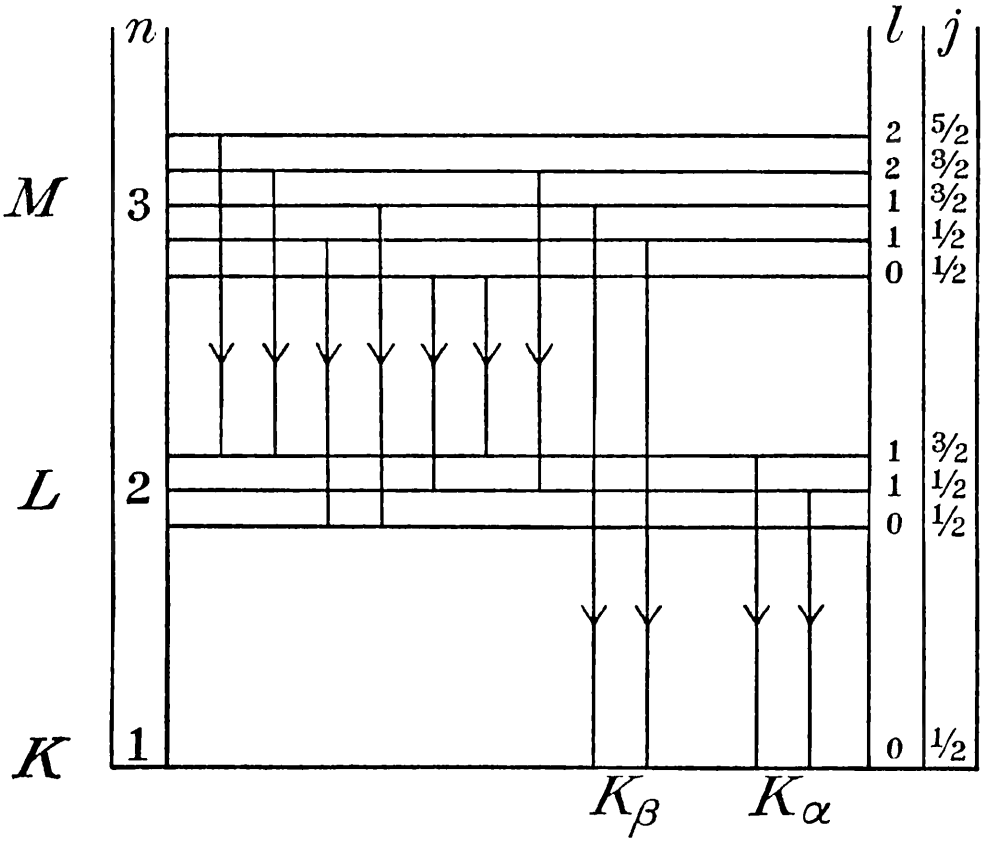

FIG. 44·4

accounted for in the same way as can that of optical spectral terms. The K term is single or in other words there is only one K energy level. There are 3 L levels, 5 M levels, and so on. The multiplicities are in fact represented by successive odd numbers. These energy levels are distinguished by the quantum numbers n, l and j in the way illustrated in Fig. 44·4.

The figure shows all the possible transitions (so far as K, L and M levels are involved) in the production of K and L lines, the selection rules being based, as our wave-mechanical studies would lead us to anticipate, on

$$\Delta l = +1 \text{ or } -1,$$
$$\Delta j = +1 \text{ or } 0 \text{ or } -1.$$

§ 44·5. The Anomalous Zeeman Phenomenon

In a sufficiently weak magnetic field the angles between j, s and l (Fig. 44·1) are not different from their values in the absence of the field and the angle between j and the field remains constant. The only effect of the field is to make j execute a ' Larmor ' precession about the axis of the field. The spectral terms whose difference, in the absence of the magnetic field **H,** gives the frequency of a spectral line such as one of the D lines of sodium, will be modified by an amount equal to the mutual energy of the electron orbit and the magnetic field divided by Planck's constant, h (cf. **40·6**). It follows from **(44·11)** that the possible values of this mutual energy are

$$j.g.BH,\ (j-1)gBH,\ \ldots$$
$$\ldots -(j-1)gBH \text{ and } -j.g.B.H$$

where B is the Bohr magneton.

Each spectral term is therefore modified by the field by the amounts

$$j.g.B.H/h,\ (j-1)g.B.H/h,\ \ldots$$
$$\ldots -(j-1).g.B.H/h,\ -j.g.B.H/h,$$

and if we represent the magnetic quantum numbers j, $j-1$, $\ldots -(j-1)$, $-j$ by the general symbol n_ϕ the frequency differences emerging from the modified spectral terms are expressed by

$$\Delta\nu = \Delta(n_\phi g).B.H/h$$

or

$$\Delta\nu = \Delta(n_\phi g)eH/4\pi m_0 c.$$

We shall illustrate the Zeeman effect which these formulae express, and which is termed the **anomalous Zeeman effect** since it differs from normal type explained by H. A. Lorentz (cf. **§§ 26·7** and **41·2**), by reference to the D lines of sodium. The line D_2 is represented by

$$3S_{1/2} - 4P_{3/2},$$

where 3 is the total quantum number, n, of the end state and 4 that of the initial state. The value of j for the initial term is 3/2 and for the final term 1/2. Hence $g = 4/3$ for the P term and $g = 2$ for the end term. The magnetic quantum numbers of the P term are

$$\frac{3}{2},\ \frac{1}{2},\ -\frac{1}{2},\ -\frac{3}{2}$$

and those of the S term are

$$\frac{1}{2},\ -\frac{1}{2}.$$

The permitted values of Δn_ϕ are $+1$, -1 and 0. The changes $+1$ and -1 give for $\Delta\nu$ the following possibilities

$$\Delta\nu = \frac{B.H}{h} \times \left\{1, \frac{5}{3}, -\frac{5}{3}, -1\right\}$$

or

$$\Delta\nu = \frac{eH}{4\pi m_0 c}\left\{1, \frac{5}{3}, -\frac{5}{3}, -1\right\}. \quad . \quad . \quad \textbf{(44·51)}$$

The lines represented by these values of $\Delta\nu$ show circular polarization when viewed along the line of the magnetic field. When $\Delta n_\phi = 0$ we find

$$\Delta\nu = \frac{B.H}{h} \times \left\{-\frac{1}{3}, +\frac{1}{3}\right\}$$

or

$$\Delta\nu = \frac{eH}{4\pi m_0 c}\left\{-\frac{1}{3}, +\frac{1}{3}\right\}, \quad . \quad . \quad . \quad \textbf{(44·52)}$$

and these lines are polarized in the sense that the electric vector is parallel to the field.

For D_1 we get

$$\Delta\nu = \frac{eH}{4\pi m_0 c}\left\{\frac{4}{3}, -\frac{4}{3}\right\}, \quad . \quad . \quad . \quad . \quad \textbf{(44·53)}$$

with circular polarization and

$$\Delta\nu = \frac{eH}{4\pi m_0 c}\left\{-\frac{2}{3}, +\frac{2}{3}\right\}, \quad . \quad . \quad . \quad \textbf{(44·54)}$$

for which the polarization corresponds to the electric vector being parallel to the direction of the magnetic field.

When the intensity of the magnetic field is increased the observed Zeeman pattern passes over gradually into the normal one. This is known as the **Paschen-Back effect.**

It may be noted that in the case of singlet terms, for which $s = 0$, the Landé factor, g, is equal to unity. Consequently lines which are due to transitions between pairs of singlet terms exhibit the normal Zeeman effect.

§ 44·6. The Gyromagnetic Anomaly

Any charged body in rotation possesses a magnetic moment and an associated angular momentum. A very simple calculation shows that **M**/**J**—**M** is the magnetic moment and **J** the angular momentum—is equal to $e/2m_0c$ where e is the charge on the body and m_0 is its (rest) mass. Consider an element of the

body in the form of an anchor ring—and let us suppose for the sake of simplicity that the electric density ρ and the mass density ρ' are uniform throughout the body—rotating about its axis with the velocity v (angular velocity v/r). If α be the cross-sectional area of the ring we shall have a convection current equal to $\rho\alpha v$ and consequently a magnetic moment equal to $\rho\alpha v\pi r^2/c$ in the units we have called mixed units. The associated angular momentum is $\rho'\alpha 2\pi r^2 v$ and the ratio of the two is $\rho/2\rho' c$ which is equal to

$$charge/(2 \times mass \times c).$$

From the point of view of the classical theory therefore we should expect that the ratio **M/J** of the magnetic moment of an electron orbit to its angular momentum would be $e/2m_0c$.

When an iron cylinder, free to rotate about its axis, is magnetized by a field parallel to its axis, the electron orbits must turn so that their magnetic moments are in the direction of the magnetic field, or at any rate have components in this direction. This will produce in the body of the iron an angular momentum, and since the field does not exert a couple about the axis of the iron cylinder the principle of conservation of angular momentum requires that the cylinder as a whole will acquire an equal and opposite momentum. We should expect the ratio of the magnetic moment of the cylinder to the angular momentum it acquires to have the value given above—according to classical principles at any rate. The effect was predicted by O. W. Richardson and is known as the **gyromagnetic effect.** The measurements of Chattock, Bates and others have proved the existence of the effect; but its value turns out to be twice the value originally expected, i.e. it appears to be equal to e/m_0c. This discrepancy is known as the **gyromagnetic anomaly.**

We can explain it by supposing the electrons in the iron atom which contribute to its magnetic moment are in S states so that their magnetic moments are those associated with their spin. Since each such electron has the magnetic moment $eh/4\pi m_0c$ and a spin momentum $h/4\pi$ the ratio of the two is in fact equal to e/m_0c. So once again the hypothesis of Uhlenbeck and Goudsmit appears to be verified.

BIBLIOGRAPHY

M. Planck: Zur Theorie des Gesetzes der Energieverteilung im Normalspektrum. Verh. d. Deutschen Physikalischen Ges., 2, p. 237 (1900).
—— Ueber das Gesetz der Energieverteilung im Normalspektrum. Ann. d. Physik, 4, p. 553 (1901).
P. Debye: Ann. d. Physik, 39, p. 789 (1912).
N. Bohr: Phil. Mag., 26, pp. 1, 476, 857 (1913).

W. WILSON : Phil. Mag., 29, p. 795 (1915) ; 31, p. 156 (1916). These papers contain the earliest statement of the quantum conditions and a description of electron orbits in terms of a radial and an angular quantum integer.

—— Proc. Roy. Soc., A, 102, p. 478 (1922). Deals with a generalization of the quantum conditions and its application to the Zeeman phenomenon.

A. SOMMERFELD : Ann. d. Phys., 51, p. 1 (1916).

Sir W. HAMILTON : Theory of Systems of Rays. Trans. Roy. Irish Acad., 15, p. 69 (1828).

W. WILSON : The Origin and Nature of Wave Mechanics. Sci. Progress, 32, p. 209 (1937).

L. DE BROGLIE : Ann. d. Physique, 3, p. 22 (1925).

—— Introduction to Wave Mechanics. (Methuen.)

NEWTON : Opticks. Messrs. Bell's reprint of the 4th edition contains a very excellent introduction by Professor E. T. Whittaker.

E. SCHROEDINGER : Abhandlungen zur Wellen Mechanik (Barth). This is a collection of Schroedinger's papers on wave-mechanics. An English translation is published by Blackie.

W. HEISENBERG : Zeitsch. für Physik, 33, p. 879 (1925).

M. BORN u. P. JORDAN : Zeitsch. für Physik, 34, p. 858 (1925) ; 35, p. 557 (1926). The latter paper is in collaboration with Heisenberg.

UHLENBECK u. GOUDSMIT. Naturwissenschaft, 13, p. 953 (1925).

P. A. M. DIRAC : Quantum Mechanics. (Oxford : Clarendon Press.)

H. T. FLINT : Wave Mechanics. (Methuen's Monographs.)

G. TEMPLE : General Principles of Quantum Theory. (Methuen's Monographs.)

A. SOMMERFELD : Atomban u. Spectrallinien. (Vieweg.) The supplement to this work contains a valuable account of wave and quantum dynamics.

W. WILSON and J. CATTERMOLE : The Elementary Particle, Phil. Mag., Ser. 7, Vol. XXVII, p. 84.

INDEX OF SUBJECTS

INDEX OF NAMES

Printed in Great Britain by Butler & Tanner Ltd., Frome and London